

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:24:19 ; Search time 16.6136 Seconds
(without alignments)

3048.258 Million cell updates/sec

Title: US-09-622-964-5

Perfect score: 2350

Sequence: 1 MITYTSQVANAELGSPSRLL.....QLLPSSVPEPLWAPSTSA 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pdb*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pdb*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pdb*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pdb*
- 5: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pdb*
- 6: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pdb*
- 7: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pdb*
- 8: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pdb*
- 9: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pdb*
- 10: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pdb*
- 11: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pdb*
- 12: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pdb*
- 13: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pdb*
- 14: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pdb*
- 15: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pdb*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	361	15.4	314	10	US-09-768-826-54
2	239	10.2	251	10	US-09-768-826-35
3	238	10.1	99	9	US-09-764-872-429
4	151	6.4	118	10	US-09-764-877-2006
5	137	5.8	60	10	US-09-764-877-1622
6	135	5.7	71	9	US-10-011-585A-243
7	134	5.7	65	9	US-10-106-698-6346
8	130	5.5	40	9	US-10-106-698-6318
9	130	5.5	76	9	US-10-012-896-575
10	130	5.5	76	9	US-10-012-896-888
11	130	5.5	76	9	US-09-895-793-575
12	130	5.5	76	9	US-09-895-793-888
13	130	5.5	76	9	US-09-895-814-575
14	130	5.5	76	9	US-09-895-814-888
15	130	5.5	76	10	US-09-759-143-575
16	130	5.5	76	10	US-09-759-143-888
17	130	5.5	76	10	US-09-780-669-575
18	130	5.5	76	10	US-09-780-669-888
19	130	5.5	76	10	US-09-822-827-575

20	130	5.5	76	10	US-09-822-827-888	Sequence 888, App
21	129.5	5.5	45	9	US-10-106-698-4849	Sequence 4849, App
22	129	5.5	72	9	US-09-925-299-1027	Sequence 1027, App
23	129	5.5	72	10	US-09-925-299-1027	Sequence 1027, App
24	125	5.3	47	9	US-10-106-698-4727	Sequence 4727, App
25	124.5	5.3	66	9	US-09-764-891-4710	Sequence 4710, App
26	121	5.1	60	9	US-09-764-872-376	Sequence 376, App
27	120	5.1	104	9	US-10-073-961-260	Sequence 260, App
28	120	5.1	104	10	US-09-764-887-260	Sequence 260, App
29	120	5.1	123	9	US-10-092-154-509	Sequence 509, App
30	120	5.1	123	10	US-09-764-847-509	Sequence 509, App
31	119	5.1	68	10	US-09-925-300-1301	Sequence 1301, App
32	118.5	5.0	74	9	US-10-001-835-210	Sequence 210, App
33	117.5	5.0	75	9	US-10-001-835-182	Sequence 182, App
34	117.5	5.0	75	9	US-10-001-835-195	Sequence 195, App
35	117	5.0	72	9	US-10-073-961-301	Sequence 301, App
36	117	5.0	72	10	US-09-764-887-301	Sequence 301, App
37	116	4.9	40	9	US-10-106-698-5126	Sequence 5126, App
38	114.5	4.9	75	9	US-10-001-835-225	Sequence 225, App
39	114.5	4.9	287	10	US-09-799-983-4	Sequence 4, Appli
40	114.5	4.9	287	10	US-09-799-983-6	Sequence 6, Appli
41	114.5	4.9	344	10	US-09-799-983-2	Sequence 2, Appli
42	113	4.8	94	9	US-10-091-504-998	Sequence 998, App
43	113	4.8	94	10	US-09-764-869-998	Sequence 998, App
44	111.5	4.7	75	9	US-10-001-835-223	Sequence 223, App
45	111	4.7	52	9	US-09-983-802-537	Sequence 537, App

ALIGNMENTS

RESULT 1
US-09-768-826-54
; Sequence 54, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PFS12PI
; CURRENT APPLICATION NUMBER: US/09/768, 826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-54

Query Match	15.4%	Score 361;	DB 10;	Length 314;
Best Local Similarity	43.6%	Pred. No. 4.2e-25;		
Matches	88;	Conservative 14;	Mismatches 78;	Indels 22; Gaps 4;
QY	172	EXLSLPHNMFWVPWFVWFANLSMKWLGGRIRDPILLOSLINEMNTLTTCGHLIAYDWTIS	231	
Db	13	ESLSKSDFNKTYWPCVWFVFTNLAAQARDGRIRDDIALCLLLELNKYRCKSMLEFHYDWTIS	72	
QY	232	IFLVYTVQVTVAVYVFFLTCLVGRQFLNP-----AKAY-----PGHE-----LDLVVVPVF	276	
Db	73	IFLVYTVQVTVAVYVFFLTCLVGRQFLNP-----AKAY-----PGHE-----LDLVVVPVF	276	
QY	277	TFLOQFFVYVGLKVLGSLRALLGWRHGQRGHGOOLLETRMOCQERKVSrv-----ESSQ	329	
Db	133	TLLOQFFVYVGLKVLGSLRALLGWRHGQRGHGOOLLETRMOCQERKVSrv-----ESSQ	329	
QY	330	AWWRTVPVTPATREAGSLEP	351	
Db	193	YVDEDPQPPYTVATAESLRP	214	

```
RESULT 2
US-09-768-826-35
; Sequence 35, Application US/09768826
; Patent No. US2002012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PFS12P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-35

Query Match          10.2%; Score 239; DB 10; Length 251;
Best Local Similarity 42.0%; Pred. No. 5.3e-14;
Matches 63; Conservative 9; Mismatches 56; Indels 22; Gaps 4;

QY 224 LYADWISIPLVYTVVTVAVYSFFLTCLVGRQFLNP-----AKAY-----PGHE----- 268
Db 2 LFHYDWISIPLVYTVVTVAVYSFFLTCLVGRQFLNP-----AKAY-----PGHE----- 268

QY 269 LBLVVPVFTLQFFVYVGMKVLGRALLGWRHGQHQQLLETRMQCOERKSVRV--- 325
Db 62 PMYVPLTLLQFFVYVGMKVLGRALLGWRHGQHQQLLETRMQCOERKSVRV--- 325

QY 326 -----ESSQAWRTVPVTPATREAGSLEP 351
Db 122 LPFAKQDQWDEDDQPPVTVATAESLRP 151

RESULT 3
US-09-764-872-429
; Sequence 429, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 429
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-872-429

Query Match          10.1%; Score 238; DB 9; Length 99;
Best Local Similarity 55.4%; Pred. No. 2e-14;
Matches 46; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFSLKLCWRGSIYKLYGFLIFLLCVYIIRPIYRLALTEEQOL 60
Db 13 MTSVYTLKVAERFGFGSLLRWRGSIYKLYGFLIFLLCVYIIRPIYRLALTEEQOL 60

QY 61 MPEKLTLCDSYIQLPIPSFVLG 83
Db 73 VYAQVARYCNRSADLPIPSFVLG 95

RESULT 4
US-09-764-877-2006
; Sequence 2006, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2006
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-2006

Query Match          6.4%; Score 151; DB 10; Length 118;
Best Local Similarity 37.3%; Pred. No. 2.5e-06;
Matches 38; Conservative 4; Mismatches 26; Indels 34; Gaps 4;

QY 329 QAWWRTVPVTPATREAGSLEPGRRLRWQSSSTPLERMMMLRPTGLSTGICRCPCW 388
Db 12 QAWWRTVPVTPATREAGSLEPGRRLRWQSSSTPLERMMMLRPTGLSTGICRCPCW 388

QY 389 LWMRCTRTCLGWSRTCTGSPSHSPPTQLLPSSVPEPLWAP 430
Db 44 ---RLHHCPTGWATECDVSVKQK-OKQLPPPTPP---WSP 79

RESULT 5
US-09-764-877-1622
; Sequence 1622, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1622
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1622

Query Match          5.8%; Score 137; DB 10; Length 60;
Best Local Similarity 68.6%; Pred. No. 2e-05;
Matches 24; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 322 VSRVSSQAWWRTVPVTPATREAGSLEPGRRL 356
Db 5 IKNIKISQAWWRTVPVTPATREAGSLEPGRRL 39

RESULT 6
US-10-011-585A-243
; Sequence 243, Application US/10011585A
; Publication No. US20030039986A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
```

```
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0261
; CURRENT APPLICATION NUMBER: US/10/011.585A
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/245,740
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 245
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 243
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-011-585A-243

Query Match      5.7%; Score 135; DB 9; Length 71;
Best Local Similarity 44.1%; Pred. No. 3.8e-05;
Matches 30; Conservative 7; Mismatches 17; Indels 14; Gaps 2;

QY 286 GMLKVLGRALLGWRHGRGHGQQLLETRMQCQKVRVSSQAWWRTVPVATREAEA 345
Db 18 GMIAGQEFNAGSGHG-----ETPSELLKIQKISR-----TWQAPVIPATQAEAA 63

QY 346 GESLEPGR 353
Db 64 GESLEPGR 71

RESULT 7
US-10-106-698-6346
; Sequence 6346, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6346
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6346

Query Match      5.7%; Score 134; DB 9; Length 65;
Best Local Similarity 54.5%; Pred. No. 4.2e-05;
Matches 30; Conservative 4; Mismatches 11; Indels 10; Gaps 2;

QY 302 GQRGHGQQLLETRMQCQKVRVSSQAWWRTVPVATREAEAGSLPGRRL 356
Db 20 GQDGETPSLLKI-----QKKISR-----AWWHVPVIPATWETEAGELLEPGRRRL 64

RESULT 8
US-10-106-698-6318
; Sequence 6318, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
```

```
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6318
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6318

Query Match      5.5%; Score 130; DB 9; Length 40;
Best Local Similarity 69.4%; Pred. No. 5.3e-05;
Matches 25; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

QY 321 KVSRRVSSQAWWRTVPVATREAEAGSLPGRRL 356
Db 6 KIQKI--SWAWWRTVPVATLEAEAGSLAPRRRL 39

RESULT 9
US-10-012-896-575
; Sequence 575, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Panger, Gary R.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-575

Query Match      5.5%; Score 130; DB 9; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 318 QERKVSRRVSSQAWWRTVPVATREAEAGSLPGRRL 355
Db 3 KSRFTKNTKITQAWRAPVTPCTREAEAGSLPGRRL 40

RESULT 10
US-10-012-896-888
```

; Sequence 888, Application US/10012896
; Publication No. US20020183251A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 888
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-012-896-888

Query Match 5.5%; Score 130; DB 9; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 318 QERKVSRYVSSQAWRTVPVTPATREAGGESLEPGRRR 355
Db 3 KSRFTKNTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

RESULT 11

; Sequence 575, Application US/09895793
; Publication No. US20020192763A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.

; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-895-793-575

Query Match 5.5%; Score 130; DB 9; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 318 QERKVSRYVSSQAWRTVPVTPATREAGGESLEPGRRR 355
Db 3 KSRFTKNTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

RESULT 12

US-09-895-793-888

; Sequence 888, Application US/09895793
; Publication No. US20020192763A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 888
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-895-793-888

Query Match 5.5%; Score 130; DB 9; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 318 QERKVSRYVSSQAWRTVPVTPATREAGGESLEPGRRR 355
Db 3 KSRFTKNTKITQAWWRAPVPGTREAEGGESLEPGRLR 40


```
RESULT 13
US-09-895-814-575
; Sequence 575, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-575

Query Match      5.5%; Score 130; DB 9; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY      318 QERKVSRESSQAWRTVPVTPATREAEAGESLEPGRRR 355
Db      3 KSRFTKTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

RESULT 14
US-09-759-143-575
; Sequence 575, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-575

Query Match      5.5%; Score 130; DB 10; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY      318 QERKVSRESSQAWRTVPVTPATREAEAGESLEPGRRR 355
Db      3 KSRFTKTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

Search completed: July 10, 2003, 12:28:57
Job time : 16.6136 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:04 ; Search time 12.7371 Seconds
(without alignments)
3283.199 Million cell updates/sec

Title: US-09-622-964-5

Perfect score: 2350

Sequence: 1 MTITVTSQVANARLGSFRL.....QLLPSSVPEPLWAPSTSA 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224/ seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	619.5	26.4	405	2	T27971
2	610.5	26.0	612	2	T32368
3	590	25.1	557	2	T32367
4	577	24.6	632	2	S43917
5	573.5	24.4	413	2	T21644
6	571	24.3	499	2	T27630
7	559	23.8	584	2	T19565
8	536	22.8	387	2	H89192
9	536	22.8	400	2	T21670
10	535	22.8	1355	2	T28715
11	529	22.5	450	2	T18781
12	528	22.5	523	2	T18782
13	512	21.8	459	2	S40708
14	500.5	21.3	420	2	B88710
15	494.5	21.0	405	2	S42371
16	491	20.9	613	2	T16885
17	476.5	20.3	512	2	T19806
18	450	19.1	513	2	T24210
19	382	16.3	400	2	T20050
20	351.5	15.0	444	2	T20048
21	347	14.8	530	2	T28037
22	331	14.1	411	2	T28038
23	319	13.6	806	2	T15468
24	285	12.1	434	2	T20922
25	113	4.8	597	4	E40201
26	104.5	4.4	457	2	D71505
27	104.5	4.4	457	2	S44484
28	104.5	4.4	613	2	G82338
29	104	4.4	516	2	A80665

30	104	4.4	673	4	P40201
31	98	4.2	741	2	F90739
32	98	4.2	741	2	H85589
33	98	4.2	786	2	H64817
34	97.5	4.1	428	2	B81531
35	97.5	4.1	432	2	D72008
36	97.5	4.1	432	2	E86616
37	97.5	4.1	553	2	T38541
38	97	4.1	681	2	S39058
39	97	4.1	716	2	H85089
40	97	4.1	735	2	AD0341
41	96.5	4.1	3010	1	A45573
42	96	4.1	785	2	A86299
43	95.5	4.1	357	2	T22879
44	95.5	4.1	418	2	S41044
45	94.5	4.0	646	2	T41545

ALIGNMENTS

RESULT 1

T27971

hypothetical protein ZK675.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000

C:Accession: T27971

R:Sim, M.

submitted to the EMBL Data Library, November 1994

A:Reference number: Z20448

A:Accession: T27971

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-405 <WIL>

A:Cross-references: EMBL:Z46812; PIDN:CAA86845.1; GSPDB:GN000020; CESP:ZK675.3

A:Experimental source: clone ZK675

C:Genetics:

A:Gene: CESP:ZK675.3

A:Map position: 2

A:Introns: 43/2; 84/2; 110/3; 174/3; 233/3; 264/2; 335/3; 372/1

C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 26.4%; Score 619.5; DB 2; Length 405;
Best Local Similarity 41.2%; Pred. No. 2.1e-46;
Matches 121; Conservative 58; Mismatches 102; Indels 13; Gaps 3;

QY 1 MTITVTSQVANARLGSFRLLLCWGSIYKLYGFLFLCYIIRFYRLALTEEQQL 60

Db 1 MTISYSD-----TFKLKLFWRKSLWKAIRWGHLLIPLTWYIINAYYRFGMTKEQQN 52

QY 61 MPEKLTLYCDSYIQLIPISFVLGVYVTLVWTRWNOYENLWPDRLMSLVSGFVGKDSQ 120

Db 53 EPIKVMVLVDGWTKEIPITFLGFIAMVRWVWVCCQLISWPDHLLNVNSALIRGQDPE 112

QY 121 GRLRLRTIIRVANLGNVILRSVSVAVKRFPSSAQHLVQAGFMTPAHKLKLSL---P 177

Db 113 TRIIRKTIARTILTSVLARSISLRVLARYPTDHDHLDVSLGMLTKEEWMFKSILVHVDP 172

QY 178 HNMFWPWWFANLSMKAWLGRIRDPDILLOSLLNEMNTLRQCGLHLYAYDMISPLVYT 237

Db 173 HQRWVPLNWTQTMWVRCFEKGTTLTHNELRVLLDALEKYNRGGFQLFYDWIAIPLVYT 232

QY 238 QWTVAVVYFPLTCLVGRQPLNPAKAYPGHELDLVVVPFTFLQFPFYGVNWKVG 291

Db 233 QVSTISVYGYFLPALIRGQY--PSKNENEELVDVTVPIFTILQFLFYGVNWKVG 284

RESULT 2

T32368

hypothetical protein C01B12.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000

C:Accession: T32368


```
RESULT 5
T21644
hypothetical protein F32B6.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21644
R:Basham, V.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19453
A:Accession: T21644
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-413 <WIL>
A:Cross-references: EMBL:Z81074; PIDN:CA803043.1; GSPDB:GN00022; CESP:F32B6.9
A:Experimental source: clone F32B6
C:Genetics:
A:Gene: CESP:F32B6.9
A:Map position: 4
A:Introns: 22/2; 61/1; 141/2; 176/3; 200/3; 290/3; 332/1

Query Match      24.4%; Score 573.5; DB 2; Length 413;
Best Local Similarity 39.9%; Pred. No. 2.3e-42;
Matches 120; Conservative 53; Mismatches 109; Indels 19; Gaps 5;

QY 1 MTITYTSQVANARLGFSRLLLCWRGSIYKLYGFIPLCYIIRFIYRLAL-----T 55
DQ 1 MTISYS-----GNVIRILLRWKGSIMRTAKKELLILYLVSVRVFLKGIDLDDDD 52
QY 56 EQOOL---MFEKLTLYCDSYIQLIPISFVLGFTVLTWTRWNNQYENLPWDRMLSLVS 111
DQ 53 EDRLLKVRMFETFCQCDSYTLIPLTLFLGFTVSVNVARMRQFTLYWPEIDLSVLC 112
QY 112 GFVEGKDEQGRLLRRLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMPAHHQL 171
DQ 113 TVLHQHDEKSKRRHTIARYNLNANLAWRDIGSKRLRPPSVHSLIESGLLTKETQIL 172
QY 172 EKLSLPH-NMFVWPVWFANLKMKGIRDPILLQSLNEMNTLRTQCGHLYAYDWI 230
DQ 173 EAMHAENESSRWITPLHWIQLMRQVEEHKPTASLFNQVGBLRFQSLRKLKLYSDWV 232
QY 231 SPLVYTVVTVAVYFFLTCLVGRQFLNPAKAYPGHGLDLVVPVFTFLOFFYVGLKV 290
DQ 233 CVELVYTVAAALATYSFFFTFLGROFLP-DIETGKELDLVVPVFTFIVQFLFVGNFKV 291
QY 291 G 291
DQ 292 G 292

RESULT 6
T27630
hypothetical protein ZC518.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000
C:Accession: T27630
R:Thomas, K.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z20396
A:Accession: T27630
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-499 <WIL>
A:Cross-references: EMBL:Z68753; PIDN:CAA92989.1; GSPDB:GN00022; CESP:ZC518.1
A:Experimental source: clone ZC518
C:Genetics:
A:Gene: CESP:ZC518.1
A:Map position: 4
A:Introns: 30/3; 60/2; 106/2; 239/3; 267/3; 315/3; 329/3; 380/3; 445/2; 466/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match      24.3%; Score 571; DB 2; Length 499;
Best Local Similarity 32.4%; Pred. No. 4.7e-42;
```

```
Matches 124; Conservative 70; Mismatches 145; Indels 44; Gaps 6;

QY 1 MTITYTSQVANARLGFSRLLLCWRGSIYKLYGFIPLCYIIRFIYRLALTEEQOL 60
DQ 1 MTISYTLDVQTNLQSPFSLLRWGSVKAVFQQLAVTAVFLISCIYRYMLSPSQOD 60
QY 61 MFEKLTLYCDSYIQL-IPISFVLGFTVLTWTRWNNQYENLPWDRMLSLVSGFVEGKDE 119
DQ 61 VFEQLIRYFDNKLDAIPLTLFLGFTVSVVVARWGSILNGIGWIDDAALLPATYIRGADE 120
QY 120 QGRLLRRLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMPAHHKLEKLSLPHN 179
DQ 121 ETRVIRNRLVYLVSQALVLRDISMQVRKRPFTMDTLAASGLMTHSEMDILDHDKDPS 180
QY 180 MFVWPVWFANLKMKGIRDPILLQSLNEMNTLRTQCGHLYAYDWISIPLVYTVQV 239
DQ 181 RYWTSIQWLSNLVYECQKKGKVDYSYLMNKIVDIDIGKFRHGLASLLKLDVVPVFLVYQV 240
QY 240 TVVAVYSFLLTCLVGRQFL-----NPAKAYPGHGLDLVVPVFTFLOFFYVGLKV----- 290
DQ 241 IFLAVRIYFIMCLIGRQFIVTGNPS-----GIDLWLPITTWQFLVYMGWKKVABALL 294
QY 291 -----GLSRALLGWRHQGHGQQLLETRMQCQERKSVRVSS 328
DQ 295 NPLGEDDDLECNYYIDKNLITGLSVDTWTKHDDTGY--SMVEEHM-----AKTPAQKKD 348
QY 329 QAWWRTVPVIPAETRAEAGESLEP 351
DQ 349 EFWGIDKIAPLYSMESAESRVHP 371

RESULT 7
T19565
hypothetical protein C29F4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T19565
R:Kershaw, J.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19143
A:Accession: T19565
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-584 <WIL>
A:Cross-references: EMBL:Z68335; PIDN:CAA92730.1; GSPDB:GN00022; CESP:C29F4.2
A:Experimental source: clone C29F4
C:Genetics:
A:Gene: CESP:C29F4.2
A:Map position: 4
A:Introns: 36/2; 78/2; 104/2; 149/2; 174/2; 221/2; 256/3; 333/3; 412/2; 484/2; 526/1;
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match      23.8%; Score 559; DB 2; Length 584;
Best Local Similarity 36.9%; Pred. No. 6.3e-41;
Matches 101; Conservative 64; Mismatches 109; Indels 0; Gaps 0;

QY 19 RLLLCWRGSIYKLYGFIPLCYIIRFIYRLALTEEQOLMFEKLTLYCDSYIQLPI 78
DQ 63 KLIFKWKGLWKAIYLDLIVWCFCYAFISVIVYALDRSQDDTFERFMQFCNRLDYPI 122
QY 79 SFVLGFTVLTWTRWNNQYENLPWDRMLSLVSGFVEGKDEQGRLLRRLIRYANLGNVL 138
DQ 123 NFMLGFTVTVINRWMTQFANLGMIDNIALFTSYMLSGNDERGRILRSIVRMCVMSQTM 182
QY 139 ILRSVSTAVYKRPSPSAQHLVQAGFMPAHHKLEKLSLPHNMFVWPVWFANLKMKAWL 198
DQ 183 VFRDIHIGVRKRPFTLETVAAGITSSSELKCYNEVESRYAKYWLGNFNTFLNLEARRE 242
QY 199 GRIRDPILLQSLNEMNTLRTQCGHLYAYDWISIPLVYTVQVTVAVYSFLLTCLVGRQFL 258
DQ 243 GRIESAYTONATABEIRTFRSGLSLTIWTVDPILMYPOLVFMFATHCYCYLVCLVSRQV 302
QY 259 NPAKAYPGHGLDLVVPVFTFLOFFYVGLKVGL 292
```

RESULT 10
T28715

Query Match

RESULT 11

Query Match 22.8%; Score 536; DB 2; Length 400;
Best Local Similarity 37.1%; Pred. NO. 4.2e-39;
Matches 108; Conservative 58; Mismatches 113; Indels 12; Gaps 4;

T18781
 hypothetical protein B0564.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T18781
 R:Lightning, J.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z19021
 A:Accession: T18781
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-450 <MIL>
 A:Cross-references: EMBL:Z73422; PIDN:CAA97765.1; GSPDB:GN00022; CESP:B0564.3
 A:Experimental source: clone B0564
 C:Genetics:
 A:Gene: CESP:B0564.3
 A:Map position: 4
 A:Introns: 60/2; 105/2; 320/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 22.5%; Score 529; DB 2; Length 450;
 Best Local Similarity 37.3%; Pred. No. 2e-38;
 Matches 110; Conservative 53; Mismatches 126; Indels 6; Gaps 2;

QY 1 MTITTSQVANARLGSFSLLLCWGSIYKLLYGEFLIFLLCYIIIRFYRLALTEEQQL 60
 DB 1 MTINYHKEIKTSHPTWFFLLPKWGSIKWAVYMETIIFLICYSIIIVYKTAMGESSQR 60

QY 61 MFEKLTLYCDSYIQLIPISFVLGFIYTLVTRWNNQYENLPWDLMSLVSGFVEGKDEQ 120
 DB 61 TRESLVRYFDKLSYIPLFVGLGFIYTLVTRWNNQYENLPWDLMSLVSGFVEGKDEQ 120

QY 121 GRLLRRLTIRYANLGNVLIIRSVSTAVYKRPSPQAHLVQAGFMTPAEHKQLEKLSLPHN- 179
 DB 121 ARIYRNRIMRYCELVLQILVFRDMSMRTRRPPTMTVVAAGFNNKHELELYNSYDTKYN 180

QY 180 ----MFVWPVWVFWPANLSMKAWLGGRIRDPILLQSLNEMNTLRQCGHLYAYDWISPLV 235
 DB 181 KLGTKIWIIPANWALCMYKARKDGYIESDYFKAQMEGEIRTWRTNIWVCNDYDWPLPLM 240

QY 236 YTVQVTVAVYSFPLTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOFFFYVGLKLV 290
 DB 241 YPQLVCLAVNLVFLVSIIRQ-LVIEKHKNVDEVDVYFPVMTFLOFFFYVGLKLV 294

RESULT 12
 T18782
 hypothetical protein B0564.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T18782
 R:Lightning, J.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z19021
 A:Accession: T18782
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-523 <MIL>
 A:Cross-references: EMBL:Z73422; PIDN:CAA97766.1; GSPDB:GN00022; CESP:B0564.4
 A:Experimental source: clone B0564
 C:Genetics:
 A:Gene: CESP:B0564.4
 A:Map position: 4
 A:Introns: 60/2; 105/2; 320/3; 405/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 22.5%; Score 528; DB 2; Length 523;
 Best Local Similarity 37.6%; Pred. No. 2.9e-38;
 Matches 111; Conservative 52; Mismatches 126; Indels 6; Gaps 2;

QY 1 MTITTSQVANARLGSFSLLLCWGSIYKLLYGEFLIFLLCYIIIRFYRLALTEEQQL 60
 DB 1 MTINYHKEIKTSHPTWFFLLPKWGSIKWAVYMETIIFLICYSIIIVYKTAMGESSQR 60

QY 61 MFEKLTLYCDSYIQLIPISFVLGFIYTLVTRWNNQYENLPWDLMSLVSGFVEGKDEQ 120
 DB 61 TRESLVRYFDKLSYIPLFVGLGFIYTLVTRWNNQYENLPWDLMSLVSGFVEGKDEQ 120

QY 121 GRLLRRLTIRYANLGNVLIIRSVSTAVYKRPSPQAHLVQAGFMTPAEHKQLEKLSLPHN- 179
 DB 121 ARIYRNRIMRYCELVLQILVFRDMSMRTRRPPTMTVVAAGFNNKHELELYNSYDTKYN 180

QY 180 ----MFVWPVWVFWPANLSMKAWLGGRIRDPILLQSLNEMNTLRQCGHLYAYDWISPLV 235
 DB 181 KLGTKIWIIPANWALCMYKARKDGYIESDYFKAQMEGEIRTWRTNIWVCNDYDWPLPLM 240

QY 236 YTVQVTVAVYSFPLTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOFFFYVGLKLV 290
 DB 241 YPQLVCLAVNLVFLVSIIRQ-LVIEKHKNVDEVDVYFPVMTFLOFFFYVGLKLV 294

RESULT 13
 S40708
 hypothetical protein C07A9.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Aug-2000
 C:Accession: S40708
 R:Smith, M.
 submitted to the EMBL Data Library, December 1993
 A:Reference number: S40701
 A:Accession: S40708
 A:Molecule type: DNA
 A:Residues: 1-459 <SMI>
 A:Cross-references: EMBL:Z29094; NID:g436440; PID:g436448
 C:Genetics:
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.8%; Score 512; DB 2; Length 459;
 Best Local Similarity 31.9%; Pred. No. 6.3e-37;
 Matches 95; Conservative 71; Mismatches 126; Indels 6; Gaps 1;

QY 1 MTITTSQVANARLGSFSLLLCWGSIYKLLYGEFLIFLLCYIIIRFYRLALTEEQ- 59
 DB 44 LSNVYNDIATSKSLMIVRMIFKRWGSLQVAKELIWIICAYLSVSVIRFALTRSQKE 103

QY 60 ----LMFEKLTLYCDSYIQLIPISFVLGFIYTLVTRWNNQYENLPWDLMSLVSGFV 114
 DB 104 QNKEIIIFERFGCYCDARMGYLPLNFVLGFFCNIIIRRLKLYSLGNIDNIALFVSAYV 163

QY 115 EGKDEQRLRLTIRYANLGNVLIIRSVSTAVYKRPSPQAHLVQAGFMTPAEHKQLEKL 174
 DB 164 RGTDDRARQIRNRIRYCVISQCLVFRDIHVGVRRRPPTLEAVAQAGIMLPHELEKFSI 223

QY 175 SLPHNMFVWPVFWPANLSMKAWLGGRIRDPILLQSLNEMNTLRQCGHLYAYDWISPL 234
 DB 224 KRIYQYVWVFWPANLSMKAWLGGRIRDPILLQSLNEMNTLRQCGHLYAYDWISPL 283

QY 235 YTVQVTVAVYSFPLTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOFFFYVGLKLV 292
 DB 284 MYPQLVNMVAVHYFFLCITRQFFISADAHNKTEVDLYIFPMTIIEIFIFNGMLKVM 341

RESULT 14
 B88710
 protein C43G2.4 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 01-Mar-2002
 C:Accession: B88710
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.el
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
 A:Accession: B88710

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:chr_IV; PIDN:AB091111.1; PID:g1572760; GSPDB:GN00022; CESP:C43G2.
C:Genetics:
A:Gene: C43G2.4
A:Map position: 4
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.3%; Score 500.5; DB 2; Length 420;
Best Local Similarity 29.2%; Pred. No. 5.7e-36;
Matches 131; Conservative 71; Mismatches 125; Indels 121; Gaps 15;

QY 1 MTITYTSQVANARLGSFRLLLCWRSIYKLYGELIFLLCYIIRF----- 48
DB 1 MTISYS-----GNFFRLLLAWKSIWRSVRELFLFLFYIRFSAPHFNYTDPT 52
QY 49 -----IYRLAITEQOQLMFEKLTLYCDSYIQLIPISFVLGPFYVTLVTRWNYENLP 101
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 53 DSKGYRKIFKVMCNFHE-----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 102 WPDRLMSLVSGFVEGDEQGLRRLRTLRVANLGNVILRSVSTAVYKRPSPSAOHLVQAG 161
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 101 WPDRLMSLVSGFVEGDEQGLRRLRTLRVANLGNVILRSVSTAVYKRPSPSAOHLVQAG 161
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 162 FMTPAEHKQLEKLSLPHNMFV--PWVWFANL-SMKAWLGGRIKDP-LIQSL----- 211
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 159 LLTEKEYEKLDINSVKAVILMLPVVTFNLCETSPSGIRWLTLHWVQQLIDAEITA 218
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 212 -----NEMNTLRQCCHLYAYDMISPLVYQVTVVAVYSFFLTCLVGRQFLN 259
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 219 GRGSVNVSVATNELKAYRISFRLLYCHDWCVPLVYVTVQALATYSYFFCLFGRQDLN 278
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 260 PAKAYPGHELDLVVVPVTFLOFFFYVGLKVG-----LSRALLG-- 298
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 279 HDQFY---SLDAFFLFTVQVFFVGVKQDLMRPFGLDDDDFELSILDRNIVTSF 335
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 299 -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 336 TIVDSIQDDDDPPKFEEDVFWKH---HNEQ---QQHQSMFLPVPTSLKNGR---IDL 384
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 340 TREAEAGESLEPGRRLLWQSSSTPLE 367
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 385 SRNAHK---HPPKLTYLEMKNQDPEE 408
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100

RESULT 15
S42371
hypothetical protein T20G5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S42371
R:Smith, A.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42368
A:Accession: S42371
A:Molecule type: DNA
A:Residues: 1-405 <SMI>
A:Cross-references: EMBL:Z30423; NID:G458479; PID:G458483
A:Genetics:
A:Introns: 23/2; 105/2; 267/3; 315/3; 366/2; 396/2; 404/1
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.0%; Score 494.5; DB 2; Length 405;
Best Local Similarity 30.2%; Pred. No. 1.8e-35;
Matches 109; Conservative 74; Mismatches 155; Indels 23; Gaps 6;

QY 1 MTITYTSQVANARLGSFRLLLCWRSIYKLYGELIFLLCYIIRFIRYRLALTEEQQL 60
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100
QY 61 MFEKLTLYCDSYIQLIPISFVLGPFYVTLVTRWNYENLPWPDRLMSLVSGFVEGKDEQ 120
DB -----YRMIPLTLLGPFVSNVSRMRQPETLR 100

Db 61 IFVRLVDVYNSRMSYVPLDMMLGFFFIAGVLRFRFWLYDIIGFDINTACSTATVIRGDSR 120
QY 121 GRLLRRLTIRYANLGNVILRSVSTAVYKRPSPSAOHLVQAGFMTPAEHKQLEKLSLPHNM 180
Db 121 AKQYRRIIRYCELTQVLI FRDLSMKARKRFPFLDTVAAAAGFPMPEKAFNFDLIQYNYNK 180
QY 181 FVVPVWFANLMSKAWLGGRIKDP-LIQSLNEMNTLRQCCHLYAYDMISPLVYQVTV 240
Db 181 YELPFNWAVALYVNAKEGLIEGDIYVTVISDIKKFRTGLAWVCNDVWVPLEIITYPTIV 240
QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVVPVTFLOFFFYVGLKVGLSRALLG-W 299
Db 241 CLAVHMYFFVGVILARQYVKGSDIP-DMIDLVPFPMTSIQFVYMGWLKVG--EGLNPNW 297
QY 300 RHQQRCHGQQLLETRMQOER-----KVSRSVSSQAWWRTPVIPATREABAG 346
Db 298 -----GEDPDDPETNNMLIDRNAMGLKIYDEGYDKTPRIE-KDAFWDDTWVPLYSEASAH 351
QY 347 E 347
Db 352 E 352

Search completed: July 10, 2003, 12:27:10
Job time : 14.0704 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:20:43 ; Search time 7.19924 Seconds
(without alignments)
2506.127 Million cell updates/sec

Title: US-09-622-964-5

Perfect score: 2350

Sequence: 1 MTITTSQVAVARLGSFRL.....QLLPSSVEPPLWAPPSTSA 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1541	65.6	585	1 VMD2 HUMAN	O76090 homo sapien
2	619.5	26.4	405	1 YS63 CAEEL	O09379 caenorhabdi
3	577	24.6	632	1 Y022 CAEEL	O45435 caenorhabdi
4	573.5	24.4	413	1 Y040 CAEEL	O45435 caenorhabdi
5	571	24.3	499	1 YHDI CAEEL	O23369 caenorhabdi
6	536	22.8	400	1 YV6L CAEEL	O19978 caenorhabdi
7	529	22.5	450	1 YQ33 CAEEL	O17528 caenorhabdi
8	528	22.5	523	1 YQ34 CAEEL	O17529 caenorhabdi
9	512	21.8	459	1 YK78 CAEEL	P34319 caenorhabdi
10	500.5	21.3	420	1 YCBL CAEEL	O94175 caenorhabdi
11	494.5	21.0	456	1 YN4 CAEEL	P34577 caenorhabdi
12	491	20.9	501	1 YSV1 CAEEL	O22566 caenorhabdi
13	450	19.1	513	1 YXAK CAEEL	O21573 caenorhabdi
14	403.5	17.2	884	1 YAVK CAEEL	O17851 caenorhabdi
15	390	16.6	81	1 VMD2 MOUSE	O88670 mus musculus
16	347	14.8	602	1 YH54 CAEEL	O18303 caenorhabdi
17	331	14.1	411	1 YH55 CAEEL	O18304 caenorhabdi
18	285	12.1	434	1 YSWJ CAEEL	O45363 caenorhabdi
19	153	6.5	585	1 ALU5 HUMAN	P39192 homo sapien
20	150	6.4	591	1 ALU8 HUMAN	P39195 homo sapien
21	150	6.4	593	1 ALU7 HUMAN	P39194 homo sapien
22	149	6.3	593	1 ALU6 HUMAN	P39193 homo sapien
23	135	5.7	591	1 ALU1 HUMAN	P39188 homo sapien
24	131	5.6	587	1 ALU2 HUMAN	P39189 homo sapien
25	127.5	5.4	603	1 ALU4 HUMAN	P39191 homo sapien
26	109	4.6	587	1 ALU3 HUMAN	P39190 homo sapien
27	106	4.5	389	1 O5C DROME	O9VHQ6 drosophila
28	104.5	4.4	457	1 SECY CLTR	P22839 chlamydia t
29	104.5	4.4	613	1 YBID VIBCH	O9KV48 vibrio chol
30	98	4.2	741	1 YBIO SCOLI	P75783 escherichia
31	97	4.1	681	1 CRY1 ARATH	O43125 arabidopsis
32	96.5	4.1	3010	1 POLG_HCVUT	O00269 h genome po
33	95.5	4.1	418	1 YXY1 HUMAN	P49646 homo sapien

RESULT 1

ID	VMD2_HUMAN	STANDARD;	PRT;	585 AA.
AC	O76090; O75904;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bestrophin (Viteliform macular dystrophy protein) (TUI5B).			
GN	VMD2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANTS BMD.			
RX	MEDLINE=98367043; PubMed=9700209;			
RA	Marquardt A., Stoeck H., Passmore L.A., Kraemer F., Rivera A.,			
RA	Weber B.H.F.;			
RT	"Mutations in a novel gene, VMD2, encoding a protein of unknown			
RT	properties cause juvenile-onset vitelliform macular dystrophy (Best's			
RT	disease).";			
RL	Hum. Mol. Genet. 7:1517-1525(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND VARIANTS BMD P-6; H-85; C-93; N-227 AND E-299.			
RX	MEDLINE=98324772; PubMed=9662395;			
RA	Petrukhin K., Koisti M.J., Bakali B., Li W., Xie G., Marknell T.,			
RA	Sandgren O., Forsman K., Holmgren G., Andreasson S., Vujic M.,			
RA	Bergen A.A.B., McGarty-Dugan V., Figueroa D., Austin C.P.,			
RA	Metzker M.L., Caskey C.T., Wadellius C.;			
RT	"Identification of the gene responsible for Best macular dystrophy.";			
RL	Nat. Genet. 19:241-247(1998).			
RN	[3]			
RP	VARIANTS BMD HIS-13; CYS-93; CYS-218; ASP-300; GLU-301 AND ILE-307.			
RX	MEDLINE=99265978; PubMed=10331951;			
RA	Caldwell G.M., Kakuk L.E., Griesinger I.B., Simpson S.A., Nowak N.J.,			
RA	Small K.W., Maumenee I.H., Rosenfeld P.J., Sieving P.A., Shows T.B.,			
RA	Ayegari R.;			
RT	"Bestrophin gene mutations in patients with Best vitelliform macular			
RT	dystrophy.";			
RL	Genomics 58:98-101(1999).			
RN	[4]			
RP	VARIANTS BMD V-10; V-82; C-92; H-96; S-135; C-218; S-218 AND K-293.			
RX	MEDLINE=99265978; PubMed=10394929;			
RA	Bakali B., Marknell T., Ingvaest S., Koisti M.J., Sandgren O., Li W.,			
RA	Bergen A.A.B., Andreasson S., Rosenberg T., Petrukhin K., Wadellius C.;			
RT	"The mutation spectrum of the bestrophin protein -- functional			
RT	implications.";			
RL	Hum. Genet. 104:383-389(1999).			
RN	[5]			
RP	VARIANTS AMD/BMD, VARIANT AVMD K-146, AND VARIANT BULL'S EYE Q-119.			
RX	MEDLINE=99381534; PubMed=10453731;			
RA	Allikmets R., Seddon J.M., Bernstein P.S., Hutchinson A., Atkinson A.,			
RA	Sharma S., Gerrard B., Li W., Metzker M.L., Wadellius C., Caskey C.T.,			
RA	Dean M., Petrukhin K.;			
RT	"Evaluation of the Best disease gene in patients with age-related			
RT	macular degeneration and other maculopathies.";			

Q92777 homo sapien
Q98146 kaposi's sa
P75291 mycoplasma
O82706 salmonella
O82710 salmonella
Q9P3H1 chlamydia m
P24686 emericella
Q9G2N6 homo sapien
P77172 escherichia
O03368 bos taurus
O03611 caenorhabdi
P26663 h genome po

```

RL Hum. Genet. 104:449-453(1999).
RN [6]
RP VARIANTS BMD F-16; C-17; N-73; H-92; C-218; H-218; L-235 AND S-296.
RX MEDLINE=21138457; PubMed=11241846;
RA Marchand D., Gogat K., Boutboul S., Pequignot M., Sternberg C.,
RA Dureau P., Roche O., Uteza Y., Hache J.C., Puech B., Puech V.,
RA Dumier V., Mouillon M., Munier F.L., Schorderet D.F., Marsac C.,
RA Dufier J.L., Abitbol M.,
RT "Identification of novel VMD2 gene mutations in patients with Best
RT vitelliform macular dystrophy."
RL Hum. Mutat. 17:225-235(2001).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: DEFECTS IN VMD2 ARE THE CAUSE OF BEST MACULAR DYSTROPHY
CC (BMD); ALSO KNOWN AS VITELLIFORM MACULAR DYSTROPHY TYPE 2. BMD IS
CC AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY TYPICAL "EGG-YOLK"
CC MACULAR LESIONS DUE TO ABNORMAL ACCUMULATION OF LIPOFUSCIN WITHIN
CC AND BENEATH THE RETINAL PIGMENT EPITHELIUM CELLS. PROGRESSION OF
CC THE DISEASE LEADS TO DESTRUCTION OF THE RETINAL PIGMENT EPITHELIUM
CC AND VISION LOSS.
CC -!- DISEASE: DEFECTS IN VMD2 COULD BE THE CAUSE OF OTHER FORMS OF
CC MACULOPATHY AS BULL'S EYE MACULOPATHY, AND ADULT VITELLIFORM
CC MACULAR DEGENERATION (AMD).
CC -!- DISEASE: IN RARE CASES, DEFECTS IN VMD2 MAY INCREASE
CC SUSCEPTIBILITY TO AGE-RELATED MACULAR DEGENERATION (AMD).
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -!- DATABASE: NAME=VMD2 mutation database;
CC WWW="http://www.uni-wuerzburg.de/humangenetics/vmd2.html".
CC -!- DATABASE: NAME=Mutations of the VMD2 gene;
CC NOTE=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/vmd2mut.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF073500; AAC64926.1; --
DR EMBL; AF073491; AAC64926.1; JOINED.
DR EMBL; AF073492; AAC64926.1; JOINED.
DR EMBL; AF073493; AAC64926.1; JOINED.
DR EMBL; AF073494; AAC64926.1; JOINED.
DR EMBL; AF073495; AAC64926.1; JOINED.
DR EMBL; AF073496; AAC64926.1; JOINED.
DR EMBL; AF073497; AAC64926.1; JOINED.
DR EMBL; AF073498; AAC64926.1; JOINED.
DR EMBL; AF073499; AAC64926.1; JOINED.
DR EMBL; AF057169; AAC64343.1; --
DR EMBL; AF057170; AAC64344.1; --
DR EMBL; AF073501; AAC33766.1; --
DR Genew; HGNC:12703; VMD2.
DR MIM; 153700; --
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Alternative splicing; Disease mutation; Polymorphism; Vision.
FT VARSPLIC 231 483
FT ASQINLPGEDEDDFTTINVDNRNQVSLAVDEMHQDLPR
FT MEFDWYNKPEPPQPPYTAASQFRASFGSTFNISLNKBE
FT WEPQNEDEEDAHAGIIGRFLGQSHDHPHPRANSRTKLL
FT WPKRESLLHGLPKNHKAQKQNVGQEDNKAKLXAVDAFK
FT SAPLYORPGYSAPOPLSPPTMPFPLEP -> GLSRALLG
FT WRHGQGHGQQLLETEMQCKERKVSSESSQAWRTPTVIPA
FT TREANAGESLEGRRLRWQSSSSPTFLERMMMLRTGLST
FT GICRPFCLWMRCRTICLGWSRTCTGTSFSPSPPTQLPPS
FT SVPEFLMAPPST (IN ISOFORM 2).
FT T -> P (IN BMD AND AVMD).
FT /FTID=VAR 000830.
FT V -> A (IN BMD).
FT /FTID=VAR 000831.
FT V -> M (IN BMD).
FT /FTID=VAR 000832.
FT A -> T (IN BMD).
FT /FTID=VAR 000833.
FT A -> V (IN BMD).
FT /FTID=VAR 010468.
FT R -> H (IN BMD).
FT /FTID=VAR 010469.
FT S -> F (IN BMD).
FT /FTID=VAR 010470.
FT F -> C (IN BMD).
FT /FTID=VAR 010471.
FT L -> V (IN BMD).
FT /FTID=VAR 000834.
FT W -> C (IN BMD).
FT /FTID=VAR 000835.
FT R -> Q (IN BMD).
FT /FTID=VAR 000836.
FT R -> W (IN BMD).
FT /FTID=VAR 000837.
FT S -> R (IN BMD).
FT /FTID=VAR 000838.
FT Q -> L (IN BMD).
FT /FTID=VAR 000839.
FT L -> V.
FT /FTID=VAR 000840.
FT I -> N (IN BMD).
FT /FTID=VAR 010472.
FT L -> V (IN BMD).
FT /FTID=VAR 010473.
FT Y -> H (IN BMD).
FT /FTID=VAR 000841.
FT R -> C (IN BMD).
FT /FTID=VAR 010474.
FT R -> H (IN BMD).
FT /FTID=VAR 010475.
FT R -> S (IN BMD).
FT /FTID=VAR 000842.
FT W -> C (IN BMD).
FT /FTID=VAR 000843.
FT Q -> H (IN BMD).
FT /FTID=VAR 010476.
FT N -> K (IN BMD).
FT /FTID=VAR 000844.
FT L -> R (IN BMD).
FT /FTID=VAR 000845.
FT D -> E (IN BMD).
FT /FTID=VAR 000846.
FT E -> Q (IN BULL'S EYE MACULOPATHY).
FT /FTID=VAR 010477.
FT G -> S (IN BMD).
FT /FTID=VAR 010478.
FT R -> H (IN BMD).
FT /FTID=VAR 000847.
FT A -> K (IN AVMD).
FT /FTID=VAR 010479.
FT S -> N (IN BMD).
FT /FTID=VAR 000848.
FT T -> I (IN AMD, SPORADIC).
FT /FTID=VAR 010480.
FT R -> C (IN BMD).
FT /FTID=VAR 000849.
FT R -> H (IN BMD).
FT /FTID=VAR 010481.
FT R -> Q (IN BMD).
FT /FTID=VAR 000850.
FT R -> S (IN BMD).
FT /FTID=VAR 000851.
FT L -> M (IN BMD).
FT /FTID=VAR 000852.
FT Y -> C (IN BMD).
FT /FTID=VAR 000853.
FT Y -> N (IN BMD).
FT /FTID=VAR 000854.

```

Query Match 65.6%; Score 1541; DB 1; Length 585;
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFRLLCWGRSIIYKLLGEPFLIFLCYIIIRFYRLALTEEQOL 60
 DB 1 MTITYTSQVANARLGSFRLLCWGRSIIYKLLGEPFLIFLCYIIIRFYRLALTEEQOL 60

QY 61 MPEKLTLYCDSYIOLIPISFVLGFTVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120
 DB 61 MPEKLTLYCDSYIOLIPISFVLGFTVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120

QY 121 GRLLRRLTYRANLGNVILRSYSTAVYKFPSPAQHLVQAGFTPTAEHKOLEKLSLPHNM 180
 DB 121 GRLLRRLTYRANLGNVILRSYSTAVYKFPSPAQHLVQAGFTPTAEHKOLEKLSLPHNM 180

QY 181 FVWPWFVFWFANLGMKAWLGGRIROPILLOSLNEMNTLRTOCGHLYAYDMISIPLVTVQV 240
 DB 181 FVWPWFVFWFANLGMKAWLGGRIROPILLOSLNEMNTLRTOCGHLYAYDMISIPLVTVQV 240

QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLKV 290
 DB 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLKV 290

RESULT 2
 YS63 CAEEL
 ID YS63 CAEEL STANDARD; PRT; 405 AA.
 AC Q09379;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK675.3 in chromosome II.
 GN ZK675.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Sime M.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z46812; CAA86845.1;
 CC WormPep; ZK675.3; CE01722.
 CC InterPro; IPR000615; Worm_fam_8.
 CC Pfam; PF01062; DUF289; 1.
 CC ProDom; PD002802; Worm_fam_8; 1.
 CC KW Hypothetical protein.
 CC SEQUENCE 405 AA; 47835 MW; 0F589D874E9E0E61 CRC64;

Query Match 26.4%; Score 619.5; DB 1; Length 405;
 Best Local Similarity 41.2%; Pred. No. 1.4e-44;
 Matches 121; Conservative 58; Mismatches 102; Indels 13; Gaps 3;

QY 1 MTITYTSQVANARLGSFRLLCWGRSIIYKLLGEPFLIFLCYIIIRFYRLALTEEQOL 60
 DB 1 MTITYSD-----TFKLFRMGSLKAWIKHLLIFLTYIINAYYRFGMTKEQON 52

QY 61 MPEKLTLYCDSYIOLIPISFVLGFTVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120
 DB 53 EFIKYVLMVDGWTKEIPITFLGFTVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 112

Query Match 24.6%; Score 577; DB 1; Length 632;
 Best Local Similarity 37.9%; Pred. No. 8.3e-41;
 Matches 110; Conservative 65; Mismatches 115; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFRLLCWGRSIIYKLLGEPFLIFLCYIIIRFYRLALTEEQOL 60
 DB 1 MTINYINAVSTSKWTWTFKLKLLKWRGSIWKAIVLELAVMLVLYGLSVIYRTALNPGQOR 60

QY 61 MFEKLTLYCDSYIOLIPISFVLGFTVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120

QY 121 GRLLRRLTYRANLGNVILRSYSTAVYKFPSPAQHLVQAGFTPTAEHKOLEKLSLPHNM 177
 DB 121 TRIIRXTIARTYTLTSLVLAWSISLVLARYPTDDHLVDLSGLTKEMVWFKSLVHVP 172

QY 178 HNMFWPWFVFWFANLGMKAWLGGRIROPILLOSLNEMNTLRTOCGHLYAYDMISIPLVTV 237
 DB 173 HQKWWVPLNWIQTMMVRCFEKGTLTHTNELRVLLDALEKYRNGFFQLFIYDIAIPLVTV 232

QY 238 QVTVVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLKV 291
 DB 233 QVSTISYGYFLPAGLIGQY--PSKNENBEIVDVYVPIETILOFLFYVGLKV 284

RESULT 3
 Y022 CAEEL
 ID Y022 CAEEL STANDARD; PRT; 632 AA.
 AC P34672;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK688.2 in chromosome III.
 GN ZK688.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsley T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R., Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";
 RL Nature 368:32-38 (1994).
 CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L16621; AAA28228.1;
 CC PIR; S44917; S44917.
 CC WormPep; ZK688.2; CE00460.
 CC InterPro; IPR000615; Worm_fam_8.
 CC Pfam; PF01062; DUF289; 1.
 CC ProDom; PD002802; Worm_fam_8; 1.
 CC KW Hypothetical protein.
 CC SEQUENCE 632 AA; 73829 MW; 0ABDD1755EF11642 CRC64;

```

Db 61 TPERIVQVCDLSVPLNFMGLFFVTVVNWYLYQIIGFDNIGLMAAEVVRGRTEQ 120
Qy 121 GRLRLRRLIRYANLGNVLILRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKQLEKLSLPHNM 180
Db 121 ARMYERNIVRYCELAQVLFDRISMRTRFRFTLDTVAAGFMFMPEHEDFDEIQYKYSK 180
Qy 181 FWVPWVFANLSMKAWLGGRIKRDPIQLQSLLEMTLRTCCGHLVAYDWISIPLVVTVQV 240
Db 181 YVFPQWAFSLTYEARKKGLIESDYQVQVQDEIKKFRIGLAWICNDYDVPPIPIYQV 240
Qy 241 TVAVYSFELTCLVGRQFLNPAKAYCPGHEDLVVVPFTFLOFFFYGVWLKV 290
Db 241 CLAVHTYELVCLLQRYVYVSEHADNKTETDLYFPIMSTLQFIYFGWGMKV 290

RESULT 4
YV4Q_CABEEL STANDARD; PRT; 413 AA.
AC O45435;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Hypothetical protein F32B6.9 in chromosome IV.
GN F32B6.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Baeham V.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z81074; CAB03043.1;
CC WormPep; F32B6.9; CE09864.
CC InterPro; IPR000615; Worm_fam_8.
CC Pfam; PF01062; DUF289; 1.
CC ProDom; PD002802; Worm_fam_8; 1.
CC Hypothetical protein.
CC SEQUENCE 413 AA; 48965 MW; A66E69A83C78790B CRC64;

Query Match 24.4%; Score 573.5; DB 1; Length 413;
Best Local Similarity 39.9%; Pred. No. 1e-40;
Matches 120; Conservative 53; Mismatches 109; Indels 19; Gaps 5;

Qy 1 MTITYTSQVANARLGSFSLLLCWRGSIYKLYGELIFLLCYIIRFYRLAL-----T 55
Db 1 MTISYS-----GNVTRILRLWKGSIWRTAKLELLIYLYSVRVFLKGIIDLIDD 52

Qy 56 EQQL-----MPEKLTLYCDSYIQLIPISFVLGYVTVVTRWNOYENLPWDRMLSLVS 111
Db 53 EDDLKMRMPFTFCQCDYSYRLPLTLTLLGFFYSNVVAVRWQFETLYWPELILSVLC 112

Qy 112 GFVEGKDEGRLLRLIRYANLGNVLILRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKOL 171
Db 113 TVLHQHDEKSKRRRTIARYLNLALAWRDISSKIRLFPSPVSHSLIESGLLTKKEYQIL 172

Qy 172 EKLSLPH-NMFPVWVFWFANLSMKAWLGGRIKRDPIQLQSLLEMTLRTCCGHLVAYDWI 230
Db 173 EAMHAENESSRITPLHWIQLIMRQVEEHKFTASLNFQVGBELRIFROSLRKLYSYDW 232

Qy 231 SIPLVTVTVAVYSFELTCLVGRQFLNPAKAYCPGHEDLVVVPFTFLOFFFYGVWLKV 290

Db 233 CVPLVTVQVAALATYSFFFTLGRQLFP-DIETGKELDLVVPVFTVQVLFVFGWFKV 291
Qy 291 G 291
Db 292 G 292

RESULT 5
YHDI_CABEEL STANDARD; PRT; 499 AA.
AC Q23369;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZC518.1 in chromosome IV.
GN ZC518.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Thomas K.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z68753; CAA92989.1;
CC WormPep; ZC518.1; CE06601.
CC InterPro; IPR000615; Worm_fam_8.
CC Pfam; PF01062; DUF289; 1.
CC ProDom; PD002802; Worm_fam_8; 1.
CC Hypothetical protein.
CC SEQUENCE 499 AA; 57229 MW; 4E87725437A5C9DF CRC64;

Query Match 24.3%; Score 571; DB 1; Length 499;
Best Local Similarity 32.4%; Pred. No. 2e-40;
Matches 124; Conservative 70; Mismatches 145; Indels 44; Gaps 6;

Qy 1 MTITYTSQVANARLGSFSLLLCWRGSIYKLYGELIFLLCYIIRFYRLALTEEQQL 60
Db 1 MTISYTLDSQTNLQSFSLLLWRGSIYKLYGELIFLLCYIIRFYRLALTEEQQL 60

Qy 61 MPEKLTLYCDSYIQL-IPISFVLGYVTVVTRWNOYENLPWDRMLSLVSGFVGKDE 119
Db 61 VPEQLIRYFDNKLDANILPTLLGFFSVFVAVRWGSIILNGICWIDDASLLPATYIRGADE 120

Qy 120 QGRLRLRRLIRYANLGNVLILRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKQLEKLSLPHN 179
Db 121 ETRVIRNLRLVLYLSQALVLRDISQVNRKRPFTMDTLAASGLMTHHEMDILDHIDKDPYS 180

Qy 180 MFVFPWVFWFANLSMKAWLGGRIKRDPIQLQSLLEMTLRTCCGHLVAYDWISIPLVYTVQV 239
Db 181 RYMTSIQNSLNLVYECQKKGVDSYLYLNKNIVDEIGKFRHGLASLLKYDWVPVLYVPOV 240

Qy 240 VTVAVYSFELTCLVGRQFL-----NPAKAYCPGHEDLVVVPFTFLOFFFYGVWLKV----- 290
Db 241 IFLAVIRYFMICLIGRQFTVTPGNS-----GIDLWLPITTMVQVFLVYGMWMAEALL 294

Qy 291 -----GLSRALLGWRHGGQGHGQQLLETRMCCQCKERVSVBSS 328
Db 295 NPLGEBDDDDLECNYYIDKNLITGLSIVDTMWKHDDTGY---SMVEEHM-----AKTPAQKDD 348

Qy 329 QAWRTPVIPATREAEAGESLEP 351
```

```

Db      349 EFWGIDKIAPLYSMESASERSVHP 371
RESULT 6
ID      YV6L CAEEL STANDARD; PRT; 400 AA.
AC      Q19578;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DE      Hypothetical protein F32G8.4 in chromosome V.
GN      F32G8.4.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      McMurray A.;
RL      Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      REVISIONS.
RA      Durbin R.;
RL      Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z72509; CAA96648.2; -
WormPep; F32G8.4; CE23700.
InterPro; IPR000615; Worm_fam_8.
Pfam; PF01062; DUF289; 1.
ProDom; PD002802; Worm_fam_8; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 400 AA; 46750 MW; B1E8ABEC3B862E84 CRC64;
-----
Query Match 22.8%; Score 536; DB 1; Length 400;
Best Local Similarity 37.1%; Pred. No. 1.4e-37;
Matches 108; Conservative 58; Mismatches 113; Indels 12; Gaps 4;
QY      1 MTITVTSQVANARLGSFSLLLCWGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
DB      1 MTISYDEB-----FSLMLRWGSIWKAIVLKOLIGFYIYIVLAFQWYLLDEKGE 52
QY      61 MFEKLTLYCDSYIQLIPISFVLGFTVTLVTRWNNQYENLPWDRMLSLVSGFVEGKDEQ 120
DB      53 YFTGWIMWCEIGAQYIPLSFLGFTVSLIVARWQFNCISIPDRKMINVSAICLFGNNM 112
QY      121 GRLLRRLIRYANLGNLIRSVSTAVYKRFPSAQLHVOAGFMTPAEHKQLEKLSLPHNM 180
DB      113 --VVRQTIRARSSLOQAIAWSGVSKLTKRPPTERRHWASKLWTEEDYLVNMTDAPHGK 170
QY      181 FWVPWVFWFANLSMKAWLGGRIKDPILLQSLNEMNTLTQCHLYAYDWISIPYTVQV 240
DB      171 WFIPLMVLNKKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 229
QY      241 TVAVYSFELTCLVGRQFLNPAKAYPGHEDLVVPVFTFLQPPFYVGVGLKVG 291
DB      230 AIATYGVFFICLIGQPKLDQDSME-KEITILFFPIFTFQMLFYLGWGLKVG 279
-----
Query Match 22.5%; Score 529; DB 1; Length 450;
Best Local Similarity 37.3%; Pred. No. 6e-37;
Matches 110; Conservative 53; Mismatches 126; Indels 6; Gaps 2;
QY      1 MTITVTSQVANARLGSFSLLLCWGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
DB      1 MTINYHKEIMTSHPTFFLLPKWKGSIWKAIVMETIIFLCYIGIISVIYKTAGESSQR 60
QY      61 MFEKLTLYCDSYIQLIPISFVLGFTVTLVTRWNNQYENLPWDRMLSLVSGFVEGKDEQ 120
DB      61 TFEISLVRYFDKRLSYIPLFELVIGFTVTVNRWTKLYQYIGFDNVLGVMANCYIRGATEK 120
QY      121 GRLLRRLIRYANLGNLIRSVSTAVYKRFPSAQLHVOAGFMTPAEHKQLEKLSLPHN- 179
DB      121 ARIYRNINRYCYELVQLVFRDMSMTRRRPFMTVVAAGFNKHELELYNSYDTKYS 180
QY      180 -----MFVPWVFWFANLSMKAWLGGRIKDPILLQSLNEMNTLTQCHLYAYDWISIPV 235
DB      181 KLGTKYWIPANWALCWTKYKARDGVIESDYFKAQMEGEIRTWRTNIEWVCNVDVPLPLM 240
QY      236 YQVTVTVAVYSFELTCLVGRQFLNPAKAYPGHEDLVVPVFTFLQPPFYVGVGLKVG 290
DB      241 YPOLVCLAVNLVFLVLSIIARQ-LVIEKHQWVDEVDVYFFVMTFLQPIFYNGWGLKV 294
-----
RESULT 8
QY      YQ64 CAEEL
ID      YQ64 CAEEL STANDARD; PRT; 523 AA.
AC      Q17529;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DE      Hypothetical protein B0564.3 in chromosome IV.
GN      B0564.3.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      Lightning J.;
RL      Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z73422; CAA97765.1; -
WormPep; B0564.3; CE05177.
InterPro; IPR000615; Worm_fam_8.
Pfam; PF01062; DUF289; 1.
ProDom; PD002802; Worm_fam_8; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 450 AA; 53275 MW; A0FED9A476166AD7 CRC64;
-----
Query Match 22.5%; Score 529; DB 1; Length 450;
Best Local Similarity 37.3%; Pred. No. 6e-37;
Matches 110; Conservative 53; Mismatches 126; Indels 6; Gaps 2;
QY      1 MTITVTSQVANARLGSFSLLLCWGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
DB      1 MTINYHKEIMTSHPTFFLLPKWKGSIWKAIVMETIIFLCYIGIISVIYKTAGESSQR 60
QY      61 MFEKLTLYCDSYIQLIPISFVLGFTVTLVTRWNNQYENLPWDRMLSLVSGFVEGKDEQ 120
DB      61 TFEISLVRYFDKRLSYIPLFELVIGFTVTVNRWTKLYQYIGFDNVLGVMANCYIRGATEK 120
QY      121 GRLLRRLIRYANLGNLIRSVSTAVYKRFPSAQLHVOAGFMTPAEHKQLEKLSLPHN- 179
DB      121 ARIYRNINRYCYELVQLVFRDMSMTRRRPFMTVVAAGFNKHELELYNSYDTKYS 180
QY      180 -----MFVPWVFWFANLSMKAWLGGRIKDPILLQSLNEMNTLTQCHLYAYDWISIPV 235
DB      181 KLGTKYWIPANWALCWTKYKARDGVIESDYFKAQMEGEIRTWRTNIEWVCNVDVPLPLM 240
QY      236 YQVTVTVAVYSFELTCLVGRQFLNPAKAYPGHEDLVVPVFTFLQPPFYVGVGLKVG 290
DB      241 YPOLVCLAVNLVFLVLSIIARQ-LVIEKHQWVDEVDVYFFVMTFLQPIFYNGWGLKV 294
-----
RESULT 7
QY      YQ63 CAEEL
ID      YQ63 CAEEL STANDARD; PRT; 450 AA.
AC      Q17528;

```

```

DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Hypothetical protein B0564.3 in chromosome IV.
GN      B0564.3.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      Lightning J.;
RL      Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z73422; CAA97765.1; -
WormPep; B0564.3; CE05177.
InterPro; IPR000615; Worm_fam_8.
Pfam; PF01062; DUF289; 1.
ProDom; PD002802; Worm_fam_8; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 450 AA; 53275 MW; A0FED9A476166AD7 CRC64;
-----
Query Match 22.5%; Score 529; DB 1; Length 450;
Best Local Similarity 37.3%; Pred. No. 6e-37;
Matches 110; Conservative 53; Mismatches 126; Indels 6; Gaps 2;
QY      1 MTITVTSQVANARLGSFSLLLCWGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
DB      1 MTINYHKEIMTSHPTFFLLPKWKGSIWKAIVMETIIFLCYIGIISVIYKTAGESSQR 60
QY      61 MFEKLTLYCDSYIQLIPISFVLGFTVTLVTRWNNQYENLPWDRMLSLVSGFVEGKDEQ 120
DB      61 TFEISLVRYFDKRLSYIPLFELVIGFTVTVNRWTKLYQYIGFDNVLGVMANCYIRGATEK 120
QY      121 GRLLRRLIRYANLGNLIRSVSTAVYKRFPSAQLHVOAGFMTPAEHKQLEKLSLPHN- 179
DB      121 ARIYRNINRYCYELVQLVFRDMSMTRRRPFMTVVAAGFNKHELELYNSYDTKYS 180
QY      180 -----MFVPWVFWFANLSMKAWLGGRIKDPILLQSLNEMNTLTQCHLYAYDWISIPV 235
DB      181 KLGTKYWIPANWALCWTKYKARDGVIESDYFKAQMEGEIRTWRTNIEWVCNVDVPLPLM 240
QY      236 YQVTVTVAVYSFELTCLVGRQFLNPAKAYPGHEDLVVPVFTFLQPPFYVGVGLKVG 290
DB      241 YPOLVCLAVNLVFLVLSIIARQ-LVIEKHQWVDEVDVYFFVMTFLQPIFYNGWGLKV 294
-----
RESULT 8
QY      YQ64 CAEEL
ID      YQ64 CAEEL STANDARD; PRT; 523 AA.
AC      Q17529;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DE      Hypothetical protein B0564.4 in chromosome IV.
GN      B0564.4.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;

```



```

CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL; U70848; AAB09111.1; -
DR WormPep; C43G2.4; CE08693.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 420 AA; 50223 MW; 5078D15E1D414A26 CRC64;

Query Match      21.3%; Score 500.5; DB 1; Length 420;
Best Local Similarity 29.2%; Pred. No. 1.4e-34;
Matches 131; Conservative 71; Mismatches 125; Indels 121; Gaps 15;

QY 1 MTITYTSQVANARLGSFSLLCWRGSIYKLLYGEFLIFLLCYIIRP----- 48
Db 1 MTISYS-----GNPFLRLKWSIWSRWFLEFLFYIRFSAPHFNYTDPT 52
QY 49 -----IYRLALTEEQLMPEKLTLYCDSYIQLIPISFVLGFTVTVTRWNOYENLP 101
Db 53 DSKGVRKLPKVCNCFHE-----YTKMIPLTFLLGFTVSNVSRWWRQPETLR 100
QY 102 WPDRLMSLVSGVEGKEQGRLLRLTRYANLGNVLILRSVSTAVYKFPFSAQHLVQAG 161
Db 101 WPEDFLSILCLLPSK--ESRPARHQIARYNLTSALAWEDYTKIRLFPFSLRNIDAG 158
QY 162 FWTPEAHKQLEKLSLPHNFWV--PWVFWANL-SMKAWLGGRIKDPF-LQSL----- 211
Db 159 LUTEYEYKQIDINYSKAVILPVPVTEKNCETSPSGIRLWTLPLHWVQQLIDAEITA 218
QY 212 -----NEMNLTQCHLYAYDWISIPLYVTVTVTVVYSPFLTCLVGHQFOLN 259
Db 219 GRGSVNYVSATNELKAYRISPRRLYCHDWCVPLVYTVQVAALATYSYFFFCFLGQDLN 278
QY 260 PAKAYPGHLDLVVPVFTFLOFFVYVGLKVG-----LSRALLG-- 298
Db 279 HDDFY---SLDAFFLFTVQVLFVGVFKVQDLMRPGLDDDDFELSYILDRNIVTSF 335
QY 299 -----WRHGQGHGQQLLETMRQCOERKSVRVSSQAWWRTVPVIPA 339
Db 336 TIVDSLQDDPPKPEEDVFWKH-----HNEQ-----QQHQSMFLPRVPTSLKNGR---IDL 384
QY 340 TREAEAGESLSPGRRLRWQSSSTPLE 367
Db 385 SRNAHK-----HPPKLTHTYLMKNQDPEE 408

RESULT 11
YX4 CAEEL
ID YX4 CAEEL STANDARD; PRT; 456 AA.
AC P34577;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein T20G5.4 in chromosome III.
GN T20G5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M., Smith A., Lloyd C.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.sib-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z30423; CAA83005.2; -
DR PIR; S42371; S42371.
DR WormPep; T20G5.4; CE28844.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 456 AA; 53136 MW; EDE22B541307F3C8 CRC64;

Query Match      21.0%; Score 494.5; DB 1; Length 456;
Best Local Similarity 30.2%; Pred. No. 4.7e-34;
Matches 109; Conservative 74; Mismatches 155; Indels 23; Gaps 6;

QY 1 MTITYTSQVANARLGSFSLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEBQOL 60
Db 1 MTVSYNQSVATSRPMTFLALIPRWGRGVSNSAIWISVWGLYFLVSATYRFLSAYQQQ 60
QY 61 MEKLTLYCDSYIQLIPISFVLGFTVTVTRWNOYENLPWDRMLSVSGFVEGKDEQ 120
Db 61 IFVRLVDVYNSRMSYVPLDMLGFFIAGVLRERFWLYDIIGFTIDNIACSTATVIRGDSR 120
QY 121 GRLRLRLTRYANLGNVLILRSVSTAVYKFPFSAQHLVQAGENTPAEHKQLEKLSLPHNM 180
Db 121 AKQYRNRIIRYCELTVQLIFRDLMSKARKRFPFLDTVAAGFMWPKHEKANFDLIQYNNK 180
QY 181 FWVPWVFWANL-SMKAWLGGRIKDPF-LQSLLEMMNTLRTQCHLYAYDWISIPLYTVTV 240
Db 181 YFLPFWAWALVYNARKEGLIEGDYVTVVISEDIKKFTGLAWVCYNDWVPLIIPITIV 240
QY 241 TVAVTSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOFFVYVGLKVLGRALLG-W 299
Db 241 CLAVHMYFPVGLARQYVKGSEIDP-DMIDLVPFPMTSIQFVFMGLKVG--EGLLPW 297
QY 300 RHGQGHGQQLLETMRQCOER-----KVSRVSSQAWWRTVPVIPATREAG 346
Db 298 -----GEDPDDFETNMLIDRLNMLGLKIVDEGYDKTPRLE-KDAFWDDTWVPLYSEASAH 351
QY 347 E 347
Db 352 E 352

RESULT 12
YX1 CAEEL
ID YX1 CAEEL STANDARD; PRT; 501 AA.
AC Q22566;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein T19C3.1 in chromosome III.
GN T19C3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U28412; AAL56625.1; -
 DR WormPep; T19C3.1; CE30182.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 501 AA; 57930 MW; 1C56C4C70707FC62 CRC64;

Query Match 20.9%; Score 491; DB 1; Length 501;
 Best Local Similarity 36.0%; Pred. No. 1e-33;
 Matches 105; Conservative 65; Mismatches 118; Indels 4; Gaps 3;

QY 1 MTITTSQVANARLGSFRLCLCWGSIYKLYGFLIFLLCYIIRFYRLALTEEQ 60
 DB 23 MTVSQQLDVSSGNPLFLRLGWRGSIWKSIVGVGDLFWLLFYAIYFYAFYAFSKQLQT 82
 QY 61 MFEKLTLYCDSYIQLIPISFVLGFTVTVTRWNOYENLPWPDRLMSLVSGFVEGKDEQ 120
 DB 83 VFEISIHDDRMKYLPLTFMLGFTVTVTRWNOYENLPWPDRLMSLVSGFVEGKDEQ 142
 QY 121 GRLLRLTRIYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180
 DB 143 DELTRAIIRYVVLHQLVFRDISMRVRRFPFLKYVDAGFMQGEQDLVLSVQESSQ 202
 QY 181 -FWPWPWFWFANLSMKAWLGGRIRDPILLOSLLNEMNTLRTQCHLYAYDWISIPLYVTQV 239
 DB 203 TYWVPINWANSALVAHQKLDQPTAFNNVIFAIKEPRVAMETLIKFDIPIAYPOV 262
 QY 240 VTVAVYSFELTCLVGRQFL-NPAKAYPGHELDLVVPTFLOFFVVGWLVK 290
 DB 263 VFLAVRVYFAICLVSRQFLSDMKST--QMDWVPVIMTVLEFIFVIGWVKV 312

RESULT 13
 YXAK CAEEL STANDARD; PRT; 513 AA.
 AC Q21973;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein R13.3 in chromosome IV.
 GN R13.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Matthews P.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z73105; CAA37442.1; -
 DR WormPep; R13.3; CE06320.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.

KW Hypothetical protein.
 SQ SEQUENCE 513 AA; 58919 MW; 666CB054AD179489 CRC64;

Query Match 19.1%; Score 450; DB 1; Length 513;
 Best Local Similarity 32.9%; Pred. No. 2.9e-30;
 Matches 100; Conservative 57; Mismatches 119; Indels 28; Gaps 5;

QY 1 MTITTSQVANARLGSFRLCLCWGSIYKLYGFLIFLLCYIIRFYR--LALTEEQ 58
 DB 1 MTVYNLDVSSASIFSLRLQLRWKSGIWKLYLKLKLFMEIIAFITVSSVYRNLIIIGET 60
 QY 59 QLMPEKLTLYCDSYIQLIPISFVLGFTVTVTRWNOYENLPWPDRLMSLVSGFVEGKD 118
 DB 61 RKINDNFAALFDQNDMDFPLTFMLGFTVTVTRWNOYENLPWPDRLMSLVSGFVEGKD 120
 QY 119 EGRLLRLTRIYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPH 178
 DB 121 DRTIRIRNRVIRYVLAQVLVPRDCSIQVRKFPPTMESIVSAGSFQCLGSSATE----- 175
 QY 179 NMFVWPWFWFANLSMKAWLGGRIRDPILLOSLLNEMNTLRTQCHLYAYDWISIPLYVTQ 238
 DB 176 -----TYWSTGLLVDAEAGKIAADLLMNEIGKHIIIEFRKMLALLSNYDWDVPIAYPO 229
 QY 239 VTVAVYSFELTCLVGRQFL-----NPAKAYPGHELDLVVPTFLOFFVVGWLVK 289
 DB 230 VFLAVRVYFAICLVSRQFLSDMKST-----TVPFVNSILOFIFVIGWVK 283
 QY 290 VGLS 293
 DB 284 VAES 287

RESULT 14
 YAVK CAEEL STANDARD; PRT; 884 AA.
 ID YAVK CAEEL
 AC Q17851;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C0989.3 in chromosome IV.
 GN C0989.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Bradshaw H.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U50069; AAB37559.2; -
 DR WormPep; C0989.3; CE29571.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 2.
 DR ProDom; PD002802; Worm_fam_8; 3.
 KW Hypothetical protein; Transmembrane; Repeat.
 FT TRANSMEM 265 285 POTENTIAL.
 FT TRANSMEM 396 418 POTENTIAL.

SQ SEQUENCE 884 AA; 102851 MW; 8DB83F9699B8FALD CRC64;
Query Match 17.2%; Score 403.5; DB 1; Length 884;
Best Local Similarity 30.8%; Pred. No. 4.2e-26;
Matches 92; Conservative 63; Mismatches 135; Indels 9; Gaps 5;
QY 1 MTTITYSOVANARLGSFRLLLCWRGSIYKLLYGEFLIFLLCYIIRFIYRLALTEQQQL 60
DB 1 MTISYTDYVATESYFGFFKFLFKWGSVWKLIIHRELFMWLVLYYTLAIYR-TLDERKK 59
QY 61 MPEKLTLYCDSYIQLIP--ISFVLGFVTVLVTWRWNOYENLWPDRLMSLVSGFEGKD 118
DB 60 IFRS---NIHFNFPSILTFMLSFPVTIVORWNVFTMGFIENAAVAVSFMK-NG 115
QY 119 EQGLRLRLIRYANIGNVLIRSVTAVYKRPFSQAHLVQAGFMTPAEHKQLEKLSLPH 178
DB 116 EDVRAQRTVIRYLVASQILWRSISIKALRRFPNYESIVTAGFLTKKESTIQTDL 175
QY 179 NMFVFWPWFANLSMKAWLGGRTRDPILOSLLNEMNTLTQCGHLYAYDWISIPLYVT 238
DB 176 DSSCVPIRAIOVLRHQYRSGNFFSHSVYRATWKEYSDFEETHLSRVKVDWVPIPLAYPQ 235
QY 239 VVTAVYSEFTLCVGRQLNPAKAVPGHELDLVVPVETFLQFFVYVGLKVLGSLRALL 297
DB 236 VIFFAVRLYFVICAFAKQIPDLDDDDARVYIHYFFIVTVFQICLGMGLKV--AEALL 292

RESULT 15
VMD2_MOUSE STANDARD; PRT; 81 AA.
AC O88870;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bestrophin (Vitellogenin macular dystrophy protein homolog) (Fragment).
GN VMD2 OR BMD1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98324772; PubMed=9662395;
RA Petrunkin K., Kolsti M.J., Bakall B., Li W., Xie G., Marknell T.,
RA Sandgren O., Foreman K., Holmgren G., Andreasson S., Vujic M.,
RA Bergen A.A., McGarty-Dugan V., Figueroa D., Austin C.P., Metsker M.L.,
RA Caskey C.T., Wadelius C.;
RT "Identification of the gene responsible for Best macular dystrophy";
RL Nat. Genet. 19:241-247(1998).
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF057171; AAC64345.1; --
DR MGD; MGI:1346332; Vmd2.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
FT NON_TER 1
FT NON_TER 81
SQ SEQUENCE 81 AA; 9413 MW; 19FD94AB3D606178 CRC64;
Query Match 16.6%; Score 390; DB 1; Length 81;
Best Local Similarity 86.4%; Pred. No. 3.9e-26;
Matches 70; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
QY 178 HMTFWPWFANLSMKAWLGGRIRDPILLOSLLNEMNTLTQCGHLYAYDWISIPLYVT 237

Db 1 HMTFWPWFANLSMKAWLGGRIRDPILLOSLLNEMNTLTQCGHLYAYDWISIPLYVT 60
QY 238 QVTVAVYSEFTLCVGRQLNPAKAVPGHELDLVVPVETFLQFFVYVGLKVLGSLRALL 297
Db 61 QVTVAVYSEFTLCVGRQLNPAKAVPGHELDLVVPVETFLQFFVYVGLKVLGSLRALL 297
Search completed: July 10, 2003, 12:24:45
Job time : 9.19924 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:21:29 ; Search time 23.8129 Seconds
(without alignments)
3763.954 Million cell updates/sec

Title: US-09-622-964-5
Perfect score: 2350
Sequence: 1 MTITTSQVANARLGSFRL.....QLLPSSVPEPLWAPPSTSA 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	843.5	35.9	465	11 Q8VCM0	Q8vcm0 mus musculu
2	836	35.6	466	4 Q9NKP0	Q9nkp0 homo sapien
3	763.5	32.5	721	5 Q9VJ36	Q9vj36 drosophila
4	692	29.4	428	6 Q8WNR7	Q8wnr7 sus scrofa
5	685	29.1	535	5 Q9VUM7	Q9vum7 drosophila
6	610.5	26.0	612	5 Q17206	Q17206 caenorhabdi
7	590	25.1	557	5 Q17205	Q17205 caenorhabdi
8	559	23.8	584	5 Q18303	Q18303 caenorhabdi
9	543.5	23.1	327	5 Q9VUM6	Q9vum6 drosophila
10	535	22.8	1447	5 Q16779	Q16779 caenorhabdi
11	493	21.0	731	5 Q9VRW4	Q9vrw4 drosophila
12	476.5	20.3	512	5 Q62095	Q62095 caenorhabdi
13	465	19.8	551	5 Q9NA59	Q9na59 caenorhabdi
14	391	16.6	551	5 Q965X4	Q965x4 caenorhabdi
15	382	16.3	400	5 Q17674	Q17674 caenorhabdi
16	351.5	15.0	444	5 Q17671	Q17671 caenorhabdi

17	181	7.7	72	4	Q9BR80	Q9br80 homo sapien
18	165.5	7.0	98	5	Q01649	Q01649 drosophila
19	134	5.7	502	4	Q9BWC6	Q9bwc6 homo sapien
20	130.5	5.6	118	4	Q13629	Q13629 homo sapien
21	127.5	5.4	827	4	Q9P225	Q9p225 homo sapien
22	122	5.2	114	6	Q9BGX7	Q9bgx7 macaca fasc
23	121.5	5.2	166	4	Q96LU6	Q96lu6 homo sapien
24	116.5	5.0	116	6	Q9N044	Q9n044 macaca fasc
25	115	4.9	52	4	Q96IL5	Q96il5 homo sapien
26	114.5	4.9	287	4	Q9BY26	Q9by26 homo sapien
27	111.5	4.7	90	6	Q9BGV8	Q9bgv8 macaca fasc
28	109.5	4.7	112	4	Q9UI61	Q9ui61 homo sapien
29	106	4.5	516	16	Q82P99	Q82p99 salmonella
30	105	4.5	68	4	Q9H389	Q9h389 homo sapien
31	105	4.5	416	2	Q9F2A1	Q9f2a1 chromatiu
32	104.5	4.4	277	4	Q96EC1	Q96ec1 homo sapien
33	104	4.4	516	16	Q82772	Q82772 salmonella
34	103.5	4.4	689	5	Q95U33	Q95u33 drosophila
35	102	4.3	151	4	Q96LU0	Q96lu0 homo sapien
36	102	4.3	328	4	Q96EG8	Q96eg8 homo sapien
37	102	4.3	3013	12	Q9QNC0	Q9qnc0 hepatitis c
38	101.5	4.3	594	8	Q95OY6	Q95oy6 tetrahymena
39	101.5	4.3	670	4	Q8TAV1	Q8tav1 homo sapien
40	101.5	4.3	695	4	Q8TCX5	Q8tcx5 homo sapien
41	100	4.3	136	4	Q96NR6	Q96nr6 homo sapien
42	99.5	4.2	3010	12	Q68533	Q68533 hepatitis c
43	99	4.2	268	4	Q9BTC5	Q9btc5 homo sapien
44	99	4.2	342	12	Q90387	Q90387 kaposi's sa
45	98	4.2	741	16	Q8X7V5	Q8x7v5 escherichia

ALIGNMENTS

RESULT 1

Q8VCM0	Q8VCM0	PRELIMINARY;	PRT;	465 AA.
ID	Q8VCM0			
AC	Q8VCM0;			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Similar to hypothetical protein FLJ20132.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Strausberg R.;			
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC019528; AAH19528.1; -			
DR	InterPro: IPR000615; Worm_fam_8.			
DR	Pfam; PF01062; DUF289; 1.			
DR	Prodom; PD002802; Worm_fam_8; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 465 AA; 51998 MW; E7145E2F893B03F6 CRC64;			

Query Match 35.9%; Score 843.5; DB 11; Length 465;
Best Local Similarity 44.1%; Pred. NO. 1.9e-65;
Matches 195; Conservative 49; Mismatches 125; Indels 73; Gaps 11;

QY	50	YRLATEEQOLMPEKLTLYCDSVIQLIPISFVLGFFVTVLVVTRWNNQYENLWPDRLMSL	109
DB	7	YFLAEEQKRYFEKLVICDQYASLIPVSFLGFTVLVHRWNNQYLCMLPDALMCI	66
QY	110	VSGFVEGDEQGRLLRRLIRYANLGNVLIRKSVSTAVYKRFPSAQHLVQAGFMTPAHK	169
DB	67	VAGTVGRDGRGLYRRLMYRVLAVLILKSVSTAVKRFPTIDHVVVEAGFMTREERK	126
QY	170	QLEKLSLPHNMFVFWFWFANLSMKAWLGGRTDIPILLQSLNEMTLRTQCGLHYADW	229
DB	127	KFENLNSYNKYWPCWFWFSSLAQAARREGRTDINSALKLLLELVNFRSKCQMLFHYDW	186

```

Qy 230 ISIPLVYQVTVAVYSPFLTCLVGRQPLNPAKAYEGHLDLVVPVFTLQFFVYVGMWK 289
Db 187 ISIPLVYQVTVAVYSPFLTCLVGRQPLNPAKAYEGHLDLVVPVFTLQFFVYVGMWK 246
Qy 290 VCLSPALLGWRHGQRGHQOQ---LLETRMQCQKRVSVESQA-----WW-----RTP 335
Db 247 V--AQQLN--PFGEDDDDFETNPLNDFQVSNLAVDEMYDDLAMLEKDYLDAAEARAP 303
Qy 336 VIPATR-----EAEAGESLEPGR-----RRLWMOSSST 364
Db 304 YTAATAPLLQPSFOGSTFDIALAKEDMQFLRDGVDGPGLVGHGDFLQRLLPAGAGSVG 363
Qy 365 PLEMMMLIR-----PTGLSTGICRCPCWLMWRCRTCLGMSRTC-----TGISPSH 411
Db 364 PLGRRLSLRLKNSCVSEASTAASCG-----CAGAADGGVGECCGDLPLDPSL 412
Qy 412 SPPTQLLPSSVEPPLWAPPST 433
Db 413 REP-ELEPPACPEPPAPPGPT 433

RESULT 2
Q9NXP0 PRELIMINARY; PRT; 466 AA.
AC Q9NXP0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CDNA FJ20132 fis, clone COL06441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Oiyashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isoigai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000139; BAA0970.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 466 AA; 52008 MW; BB9608363A09C2EC CRC64;

Query Match 35.6%; Score 836; DB 4; Length 466;
Best Local Similarity 63.5%; Pred. NO. 8.6e-65;
Matches 153; Conservative 35; Mismatches 53; Indels 0; Gaps 0;

Qy 50 YRLALTEQQLMFEKLTLYCDSYIOLIPISFVLGFFVTVLVTRWNNQYENLPKPOLMSL 109
Db 7 YRFVTEGQKRFEXKLVYCDQYASLIPVSFVLGFFVTVLVTRWNNQYENLPKPOLMSL 66
Qy 110 VSGFVEGDEQGLRLRLTIRYANLGNVLILRSVSTAVYKRPFSAGHLVQAGFMTPAEHK 169
Db 67 VAGTVHGRDGRGLYRRLTMRVYAGLSAVILRSVSTAVYKRPFTIDHVEAGFMTPEERK 126
Qy 170 QLEKLSLPHNFWPVPWFPANISMKAWLGGRTRDPILLOSILNEMNTLTQCHLYAYDW 229
Db 127 KPENLNSYNKYWPCVWFPSNLAAQARREGIRDNLSALKLLEELNVFRGKCMPLFYDW 186
Qy 230 ISIPLVYQVTVAVYSPFLTCLVGRQPLNPAKAYEGHLDLVVPVFTLQFFVYVGMWK 289
Db 187 ISIPLVYQVTVAVYSPFLTCLVGRQPLNPAKAYEGHLDLVVPVFTLQFFVYVGMWK 246
Qy 290 V 290
Db 247 V 247

```

```

RESULT 3
Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-WAY-2000 (TREMBlrel. 13, Created)
DT 01-WAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CG6264 protein (LP05915P).
GN BEST OR CG6264.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID:7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA April J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoskin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muszy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Taveanli B.C., Pappu K., Mardon G.;
RT "dbest, a Drosophila Homolog of Human Bestrophin.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003686; AAF54503.1; -.
DR EMBL; AF218817; AAF32327.1; -.
DR EMBL; AY061546; AAL29094.1; -.

```

```
DR FlyBase; FBgn0040238; best.
DR InterPro; IPR000615; Worm_fam_8
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 721 AA: 79642 MW: 8
```

Query Match	'	Score	763.5;	DB	5;	Length	721;
Best Local Similarity	49.4%;	Pred. NO.	3e-58;				
Matches 152;	Conservative	53;	Mismatches	80;			
			Indels	23;	Gaps	5;	

```

1  MTITYTQVANAR-LGSPSRLLLCWRGSIYKLLYGSEFLIFLLCYIYIRFIYRLALTBEQQ 59
      :::::
      Db
1  MTITYTEVATCRGFGGCFLLKLLRWGRSIIYKLVMLDLALFTIYAINVMYRFGLNPAQK 60
      :::::
      QY
60  LMFEKFLYCDSYIQLIPISVLGFIYTVLVTRWNNQYENLPWDRMLSLVSGFVGEKDE 119
      :::::
      Db
61  ETPEAIVQYCDSYRELPLPSVLGFIYYSIVNTRWNNQYTSIPWDPPTAFVYSSNVHCODE 120
      :::::

```

Qy	120	QGRLLRRTLI	RVANLGNVL	ILRSVSTAVYK	FFSAHLVQAGFMT	PABHKQLKL--SLP	177
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Dd	121	RGEMMRIT	IMRVCLCTM	LNVANSPVKRR	FGLNNLVGALNDNEKT	IETMNKAF	180
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	

Qy	178	H-NMFVPPVWFANLSMKAWLGGRIDRIFLLQSLNENMTLRTCQGHLYAYDWISIPVY	236
Db	181	RPSKHWLPVVAASIIITRAKEGRIRDFAVKIIIDENKFRGCGLLSIYDTISVPIY	240

Qy	237	TQWTVAVYSFELTCLVGRQ-----FLNPAKAYPGHGLDVVPVFETLQFFPYV	285
Dd	241	TQWTVAVYSFELTCMGQQQTGKVGVNTVLN-----KVDLVFPVFETLQFFPYM	292

Qy	286	GWLKVGLS	293
Db	293	GWLKVAES	300

RESULT 4
Q8WNR7

ID	Q8WNR7	PREDIMINARY;	PRT;	428 AA.
AC	Q8WNR7			
DT	01-MAR-2002	(TREMBLrel. 20, Created)		
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)		

[illegible]

mammaria; Eucheiria; Cetartiodactyla; Suina; Suidae; Sus.
 NCBI_TaxID=9823;
 [1]
 SEQUENCE FROM N.A.
 RN
 RP

Wamrostein L.I.; McLaughlin F.J., Stanton B., Yan L., Crabb J.W., Wamrostein A.D.;
 RA "Bestrophin interacts with the b-catalytic subunit of protein
 RT phosphatase 2A.";
 RT

EMBL; A1064707; AAL40882.1; -;
 DR interPro; IPR000615; Wormfam_8.
 DR Pfam; PF01062; DUF389; 1.
 DR

```

SQ      NON TER      I      FDOU280Z; WOIRL_TAM_B; I.
SEQUENCE 428 AA; 48498 MW; 3E5991B2BCDB6383 CRC64;
```

Best match: 29.4%; Score 692; DB 6; Length 428;
Best local similarity 89.7%; Pred. 0.2.9e-52;
Matches 122; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

[illegible]

274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
10

QY 275 VFTELQFFFYVGLKV 290
:|||||
Db 122 LFTLQFFFYAGLKV 137

RESULT 5						
ID	Q9VUM7	PRELIMINARY;	PRT;	535 AA.		
AC	Q9VUM7					
DT	01-MAY-2000	(TremBirel_13,	Created)			
DT	01-MAY-2000	(TremBirel_13,	Last sequence update)			
DT	01-JUN-2002	(TremBirel_21,	Last annotation update)			
DE	CG12327 protein					

GN CG12327.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Artthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 DV VERY THIN 004-0500;
 004-0500;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
na

[illegible]

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A Chandra I.,

RA de Pablo B., Carley S., Banke C., Davenport L.B., Davies P.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
PA Telatin M.

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
PA Marbulov C., Mikhlin V.

Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert V.

Shue B.C., Siden-Klamos I., Simpson M., Skupeki M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svarksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z., Weissman D.

Williams S.M., Woodgate T., Worley G.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Zhang D. A. M. *Journal of Molecular Biology* 361: 1-12 (2002)

"The genome sequence of *Drosophila melanogaster*,"
Science 287:2185-2195 (2000).
EMBL; AE003531; AAF49648.1; -.
FlyBase. FBst0035403. GenBank.

DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD02802; Worm_fam_8; 1.
SO SEQUENCE 515 AA, 52710 MDL.

Query Match 29.1%; Score 685; DB 5; Length 535;
Best Local Similarity 42.9%; Pred. No. 1.5e-51;
Matched 126; Concentration 60; Mismatch 10

1 MTITYSQVAN-ARLGSPFRLLLCWGRSIYKLLYGEFLIFLLCYIIRFIYRLALTEQQ

```

Db 1 MTVSATVATCSHFGEKWLMLRWRASIYKIIVDILAFSLCSFYEMAVIVRYALRDVDK 60
Qy 60 LMFEKLTLYCDSYIOLIPISFVLGFTVLTWTRWNOYENLPWDRMLSLVSGFVEGKDE 119
Db 61 PVFEDIVMCHSYSNLIPLSFVLGFTVLTWTRWNOYENLPWDRMLSLVSGFVEGKDE 120
Qy 120 QGRLLRRLTIIRYANLGNVLIILRSVSTAVYKRPSPSAHLVQAGFMTPAASHKOLEKLSL--- 176
Db 121 HGRLLRRLTIIRYANLGNVLIILRSVSTAVYKRPSPSAHLVQAGFMTPAASHKOLEKLSL--- 180
Qy 177 PHNMFVFWFANLGMKAWLGGRIIDPILQSLNEMNTLRTQCGHLYAYVDWISIPLVYTV 236
Db 181 KHPKYWPIVAAISIVTRAKREGRIWDPSLSKSMIDELNFKRAGCNMLIHYDTISVPLVY 240
Qy 237 TQVTVVAYSFELTCLVGRQFLNPAKAYPGHLDLWVPVFTLQFFVYVGNLKV 290
Db 241 TQVTVLAVYSYFVASIFGHOWIDRIDIKHYNINVSYYFPLFSTLEPFFFMGLKV 294

RESULT 6
ID O17206 PRELIMINARY; PRT; 612 AA.
AC O17206;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE CO1B12.3 protein.
GN CO1B12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025458; AAB70976.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 612 AA; 71031 MW; DFBB43916541DD44 CRC64;

Query Match 26.0%; Score 610.5; DB 5; Length 612;
Best Local Similarity 32.7%; Pred. No. 5.7e-45;
Matches 130; Conservative Indels 47; Gaps 6;

Qy 1 MTITYSQVANRLGFSRLLCWRGSIYKLLYCEFLIFLLCYIIRFYRLALTEBQQL 60

```

```

Db 1 MTITYSLDVASSFFCLYKLLFRWKGSIKSWAELVWLCLYAVLSVYVYVCLLTMKORA 60
Qy 61 MFEKLTLYCDSYIOLIPISFVLGFTVLTWTRWNOYENLPWDRMLSLVSGFVEGKDEQ 120
Db 61 TFEDLCIFFDTYSNFIDITFNLGFTVSAVFRWQIIFDNIGMIDTCLMIWYQIKGETER 120
Qy 121 GRLRLRRLTIIRYANLGNVLIILRSVSTAVYKRPSPSAHLVQAGFMTPAASHKOLEKLSLPHNM 180
Db 121 AKCVRNCIRYSILTOAVYRDVAASVRKRPFTFNHLVTLAGLMTKEKNAEFESIPSPHAK 180
Qy 181 FWPVFWFANLGMKAWLGGRIIDPILQSLNEMNTLRTQCGHLYAYVDWISIPLVYTV 240
Db 181 YQPMHMLFSMITLARDGMISSDIIVYDLMEKWRQFRVNLSTLFDWVPVPLVYTV 240
Qy 241 TVAVYSFELTCLVGRQFLNPAKAYPG---HELDLWVPVFTLQFFVYVGNLKVGLSRALL 297
Db 241 HLAVSFPLIALFCRQYLHPESNLNDPKQITIDLYVPIMSLLQFIFFIGMKV---AEVLL 298
Qy 298 -----GWRHGQRGHGQQLLETRMQCQERKVSRESSQAWWTFVPIPATREAEA 345
Db 299 NPLGDDDDDFECNW-----ILDRNLQ-----VGLMVVDYAYNRYPTL----- 335

RESULT 7
ID O17205 PRELIMINARY; PRT; 557 AA.
AC O17205;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE CO1B12.5 protein.
GN CO1B12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025458; AAB70976.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 557 AA; 64076 MW; 8A68BBB1436D0331 CRC64;

```

	Query Match	23.8%;	Score 559;	DB 5;	Length 584;
	Best Local Similarity	36.9%;	Pred. No. 1.7e-40;		
	Matches 101;	Conservative 64;	Mismatches 109;	Indels 0;	Gaps 0
QY	19	RULLCWGRSGIYKLLGYCEPLIPLLCYIIRFYRRLATBEOQLMPEKUTLYCDSVIQLIPI	78		
Db	63	KLIFKWKGLSKWAIYLDLITWFCYAFISVIYRYALDRSQDTEPFRWQFCNRELDYIPI	122		
QY	79	SFVLGFYVTVLVTRWNNQYENLPWDRMLMSLVGSGVEKDEQGLLRRTTIRYANLGNVL	138		

```
DR FlyBase; FBgn0036491; CG7259.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 327 AA; 38594 MW; BE7CE3BC05370F46 CRC64;

Query Match      23.1%; Score 543.5; DB 5; Length 327;
Best Local Similarity 44.3%; Pred. No. 28-39;
Matches 101; Conservative 48; Mismatches 72; Indels 7; Gaps 2;

QY 68 YCDSYLIQIPISFVLGVYTLVTRWNOYENLPWDRMLSVSGFVEKDGQGLLRRT 127
DB 3 YCKKNGSLIPLSFVLGVYTLVTRWNOYENLPWDRMLSVSGFVEKDGQGLLRRT 62

QY 128 LIRYANLGNVLIRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEK-----SLPHNMFV 182
DB 63 ILRYVCLQCVIVFTWISPRVKRFPYTTQIIEAGFLENEKKIIEAMDQAPSPYKHH--W 120

QY 183 VPWVFWANLSMKAWLGGRIIDPILQSLNENMTLTQCGHLYAYDWISIPLYVTVVTV 242
DB 121 MPVWAAISVMPARRENKIRDDYAVKTIIDELNQLRGNGCFLLYYDWISVPLVYTVVTV 180

QY 243 AVYSFELCLVQGRQFLNPAKAPYGHGLDLVVPVTFLOQFFVYVGMVKV 290
DB 181 ATYSFELSVLQQQNESHSDGTRIRRWFPILTVLQFFFTMGMLKV 228

RESULT 10
O16779 PRELIMINARY; PRT; 1447 AA.
AC O16779; O16781;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical.164.2 kDa protein.
GN T21D12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Woessner J.;
RT "The sequence of C. elegans cosmid T21D12.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016687; AAK72064.1; -.
DR HSSP; P56276; 1TLK.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003606; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_Out.
DR InterPro; IPR003591; LRR_typ.
DR InterPro; IPR000815; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.

DR Pfam; PF00047; Ig; 3.
DR Pfam; PF00560; LRR; 12.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICRPT.
DR ProDom; PD002802; Worm_fam_8; 1.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00410; IG_Like; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 13.
KW Hypothetical protein; Immunoglobulin domain.
SQ SEQUENCE 1447 AA; 164168 MW; 4BC1FE38E9E85B5 CRC64;

Query Match      22.8%; Score 535; DB 5; Length 1447;
Best Local Similarity 37.7%; Pred. No. 68-38;
Matches 115; Conservative 61; Mismatches 111; Indels 18; Gaps 5;

QY 1 MTITYTSQVANARLGSFSRLLCWRGSIYKLYGSEFLIFLCYIIRFIYR--LALTEEQ 58
DB 879 MTVSYNSDVSSSTSLIAFLRLQLRWGWSVWRELFIISFCFAITSIYRTNYFLSDEQ 938

QY 59 QLMFEKLYCDSYLIQIPISFVLGVYTLVTRWNOYENLPWDRMLSVSGFVEKDG 118
DB 939 RVFMDNFAALFDQTLNLYPLTFMLGFFVTIIVGRWNDIFLNIGWVDNTALLIATYIRGGD 998

QY 119 EQGRLRLRTLRVYANLGNVLIRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKSL-P 177
DB 999 EKSRIWRRTALRYLVLTQVLIIFRDISLVVYKGFPLTETVVASGFMLESEKRYAALSLGK 1058

QY 178 HNMFWVPVWVFWANLSMKAWLGGRIIDPILQSLNENMTLTQCGHLYAYDWISIPLYVT 237
DB 1059 YPKYFLPIQWCFSLYDARAQCKIGADYVNLKISVCGDFRKGGLQLLNDFWVPIPLVYP 1118

QY 238 QVTVVAVVSFFLTCVLRQGLF-----NPAKAPYGHGLDLVVP-VFTFLOQFFVYVGM 288
DB 1119 QVFLAVRIYFPLCLIRQSVLIDGKPKDNSPVYP-----FVPFLMTSLOQFFVYVGM 1172

QY 289 KVGLS 293
DB 1173 KVABS 1177

RESULT 11
O9VRW4 PRELIMINARY; PRT; 731 AA.
AC O9VRW4;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG10173 protein.
GN CG10173.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champs C.R., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
```


DR EMBL; Z82095; CAB05710.1; JOINED.
DR EMBL; Z82095; CAB05029.1; --
DR EMBL; Z83221; CAB05029.1; JOINED.
DR InterPro: IPR000615; Worm_fam 8.
DR Pfam: PF01062; DUF289; 1.
DR ProDom: PD002802; Worm_fam 8; 1.
SQ SEQUENCE 400 AA; 46827 MW; D1CDF601E25E2D01 CRC64;

Query Match 16.3%; Score 382; DB 5; Length 400;
Best Local Similarity 31.2%; Pred. No. 3.1e-25;
Matches 100; Conservative 53; Mismatches 113; Indels 54; Gaps 8;
QY 1 MTITYTSQVANARLGSFRLLCWRGSIYKLYGEFLPLCYIIRPIYRLAL-----T 55
Db ||||| :
1 MTITYYRAVSTESISNPLKRIIRHWHGSLAKSVWQEEFLWLPFYVVCAYRCVLP 60
QY 56 BEQQLMFEKLTLYCDSYIQLIPISFVLGFTVLTVTWNNQYENLPWPORLMSLVSGFVE 115
Db :
61 DQFKKFEILSTHQDNY---IPLEFVLGFTVTVIDRWKSFQNIPIESCAPAVSAALP 117
QY 116 GKDEQGRU-----LRRTLIYANLGNVLILRSVSTAVYKRPSPQAHLVQAGFM 163
Db ||||| :
118 G---HGALEVSAAGMDTKLTARTTIIRYLVLSQLLFREFSVKVKCRFVDRSLVDSKFL 174
QY 164 TPAEHKQL--EKLSLPHNMFWPWWWFANL-----SMKAWLGGRIKRDPIILQL 210
Db :
175 TENELTELSEELKTKRYSIILPINWAFSILRTEKLNPNQPFMNAW----- 220
QY 211 LNMENLTATQCGLYAYDWISIPLVYTVQVTVAVYSPFLTCLVGRQFLNPAKAYPCHELD 270
Db :
221 -NVINDWQVKLTLLRNGDFIPIPLAYPOAVFLAVRYFLVCLFTRQHLDS---DHHAID 276
QY 271 LVVPVFTFLQFFPYGWLKV 290
Db :
277 YFFPLLTSTFQFIYVGMKV 296

Search completed: July 10, 2003, 12:26:19
Job time : 25.8129 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:19:13 ; Search time 34.021 Seconds
(without alignments)
2158.111 Million cell updates/sec

Title: US-09-622-964-29

Perfect score: 2944

Sequence: 1 MTITYTKVANARLGSFSSL.....EHAESYPRDEAGTKPVLYE 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2944	100.0	551	20	Mouse CGICE protei
2	1899	64.5	585	20	Human CGICE long f
3	1297	44.1	435	20	Human CGICE short
4	961	32.6	721	22	Drosophila melanog
5	858.5	29.2	535	22	Drosophila melanog
6	752	25.5	327	22	Drosophila melanog
7	728	24.7	368	22	Novel human diagno
8	714.5	24.3	731	22	Drosophila melanog
9	641	21.8	314	22	Human secreted pro
10	534	18.1	251	22	Human secreted pro

11	534	18.1	251	23	ABG65506	Human albumin fusi
12	444	15.1	261	20	AAW95345	Human adult retina
13	374	12.7	101	22	ABB14768	Human nervous syst
14	313	10.6	701	22	ABG08608	Novel human diagno
15	232	7.9	99	22	AAU22652	Novel human colon
16	232	7.9	99	22	AAW92658	Human digestive sy
17	207.5	7.0	377	22	ABG05411	Novel human diagno
18	165	5.6	30	20	AAU23956	Human CGICE partia
19	122.5	4.2	188	23	ABW92888	Human polypeptide
20	122	4.1	47	22	ABG08607	Novel human diagno
21	119	4.0	30	20	AAU29968	C. elegans CGICE r
22	119	4.0	30	20	AAU29971	C. elegans CGICE r
23	115	3.9	30	20	AAU29969	C. elegans CGICE r
24	110	3.7	384	21	AAB26452	Drosophila melanog
25	110	3.7	384	22	ABB66539	Drosophila melanog
26	109.5	3.7	1816	21	AAU95440	Caenorhabditis ele
27	107.5	3.7	470	23	AAO14194	Human transporter
28	104.5	3.5	389	21	AAB20926	Drosophila odorant
29	102.5	3.5	1522	20	AAW99302	Human BA13 protein
30	102.5	3.5	1522	21	AAB23601	Human secreted pro
31	102	3.5	30	20	AAU29970	C. elegans CGICE r
32	101.5	3.4	701	22	ABG25778	Novel human diagno
33	99.5	3.4	361	23	ABP26907	Streptococcus poly
34	98.5	3.4	361	23	ABP23762	Streptococcus poly
35	99.5	3.4	855	13	AAU29815	S receptor kinase
36	99	3.4	477	21	AAG27305	Arabidopsis thalia
37	99	3.4	477	21	AAG27304	Arabidopsis thalia
38	99	3.4	533	21	AAG27303	Arabidopsis thalia
39	97	3.3	642	22	ABW9242	Drosophila melanog
40	96.5	3.3	567	22	AAB94166	Human protein sequ
41	96.5	3.3	839	20	AAU30152	A human vanilloid
42	96.5	3.3	839	20	AAU30153	A partial human va
43	96.5	3.3	839	20	AAU30155	A human vanilloid
44	96.5	3.3	839	20	AAU06558	Human capsaicin re
45	96.5	3.3	839	21	AAB332127	Human vanilloid re

ALIGNMENTS

RESULT 1

AAU29955

ID AAU29955 standard; Protein; 551 AA.

XX

AC AAU29955;

XX

DT 22-NOV-1999 (first entry)

XX

DE Mouse CGICE protein sequence.

XX

KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

KW BMD; age-related macular dystrophy.

XX

OS Mus sp.

XX

PN WO9943695-A1.

XX

PD 02-SEP-1999.

XX

PF 22-FEB-1999; 99WO-US03790.

XX

PR 25-FEB-1998; 98US-0075941.

XX

PR 18-DEC-1998; 98US-0112926.

XX

PA (MERI) MERCK & CO INC.

XX

PA (UYUP-) UNIV UPPSALA.

XX

PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;

XX

DR WPI; 1999-540560/45.

XX

DR N-FSDB; AAZ21229.

XX

PT Human and mouse polynucleotides encoding CGICE polypeptides

Db 417 RESLLHEGLPKNHKAQKQVRQEDNKAKLKAVDAPKSGPLQYRPGYYSAPQTPLSPTP 476

Qy 469 --HPTQAPSS--SDTG-----DGPSTDYQICHMKKKT 499

Db 477 MFFPLEPAGSKLHVSVTGIDTKSLKTVSSGAKSKFELLSSDGMALMEHPVSOVRRKT 536

Qy 500 VEFNL-NTPESPTEHLQRRDDOMSTNIOALMKEHAESV---PYDEEA 543

Db 537 VEFNLTDMPEIPENHLKE-PLFQSPNTNHTTUKDHMDPYVALENRDEA 583

RESULT 3

AA29954

ID AA29954 standard; Protein; 435 AA.

XX AC

XX AC

XX AC

DT 22-NOV-1999 (first entry)

XX Human CG1CE short form protein sequence.

XX CG1CE; Best's macular dystrophy; mutation; diagnosis; detection;

KW BMD; age-related macular dystrophy.

XX Homo sapiens.

OS WO9943695-A1.

PN 02-SEP-1999.

PD 22-FEB-1999; 99WO-US03790.

XX 25-FEB-1998; 98US-0075941.

PR 18-DEC-1998; 98US-0112926.

XX (MERI) MERCK & CO INC.

PA (UTUP-) UNIV UPPSALA.

XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;

PI WPI; 1999-540560/45.

XX N-PSDB; AA221228.

DR Human and mouse polynucleotides encoding CG1CE polypeptides

XX Claim 7; Fig 5; 67pp; English.

XX The present sequence represents the human CG1CE protein. When the CG1CE

CC gene is mutated it is responsible for Best's macular dystrophy (BMD).

CC Polynucleotides encoding CG1CE are useful for diagnosing whether a

CC patient carries a mutation in the CG1CE gene. Normal and mutated

CC CG1CE proteins are useful for identifying activators and/or inhibitors

CC of these proteins, in order to treat BMD. The CG1CE gene offers a

CC simpler and cheaper method of diagnosing BMD without the need for the

CC presence of the patient. The gene may also be useful to discovering

CC the genetic cause of age-related macular dystrophy.

XX SQ Sequence 435 AA;

Query Match 44.1%; Score 1297; DB 20; Length 435;

Best Local Similarity 82.4%; Pred. No. 1.3e-125;

Matches 239; Conservative 23; Mismatches 28; Indels 0; Gaps 0;

Qy 1 MTITVTNKANARLGSFSLLCWRGSIYKLYGBFLVFIFLYSIRGLYRMVLSDDQL 60

Db 1 MTITTSQVANARLGSFSLLCWRGSIYKLYGBFLVFIFLYSIRGLYRMVLSDDQL 60

Qy 61 LFEKALYCDYSYIQLIPISFVLGFFVTLVVRWMSQYENLPWDRMLIQVSSFVEGKDE 120

Db 61 MPEKLTLYCDSYIQLIPISFVLGFFVTLVVRWMSQYENLPWDRMLIQVSSFVEGKDE 120

Qy 121 GELLRTLRIRYANLGNVLLRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

Db 121 GELLRTLRIRYANLGNVLLRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

Qy 181 FWVPWVFPANLSPKAYIGGRIRDTVLQSLANVEUCLTTCQGLYAYDWISIPLYTVQV 240

Db 181 FWVPWVFPANLSPKAYIGGRIRDTVLQSLANVEUCLTTCQGLYAYDWISIPLYTVQV 240

Qy 241 TVAVTSFELACLIGRQFLNPNKDYPGHEMDLVVPVFTILQFLFYNGWLKV 290

Db 241 TVAVTSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTILQFLFYNGWLKV 290

RESULT 4

ABB62213

ID ABB62213 standard; Protein; 721 AA.

XX AC

XX AC

XX AC

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 13431.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL06316.

DR New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions

XX Disclosure; SEQ ID NO 13431; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 721 AA;

Query Match 32.6%; Score 961; DB 22; Length 721;

Best Local Similarity 50.4%; Pred. No. 2.4e-90;

Matches 191; Conservative 70; Mismatches 108; Indels 10; Gaps 6;

Qy 1 MTITVTNKANAR-LGSPSSLLLCWRGSIYKLYGBFLVFIFLYSIRGLYRMVLSDDQL 59

Db 1 MTITVTNKANAR-LGSPSSLLLCWRGSIYKLYGBFLVFIFLYSIRGLYRMVLSDDQL 59

Qy 60 LFEKALYCDYSYIQLIPISFVLGFFVTLVVRWMSQYENLPWDRMLIQVSSFVEGKDE 119

Db 61 ETEFAIVQCDYSYRELILFLSVLGVFVSVIVTRWNQVTSIPWDPDIAVFSNVHGQDE 120

XX WPI; 2001-656860/75.
 DR N-PSDB; ABL09017.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX
 PS Disclosure; SEQ ID NO 21534; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic/DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 731 AA;
 Query Match 24.3%; Score 714.5; DB 22; Length 731;
 Best Local Similarity 46.7%; Pred. No. 1.1e-64;
 Matches 142; Conservative 50; Mismatches 103; Indels 9; Gaps 4;
 QY 78 ISVLGFTVTLVSRWMSQYENLPWDRMLQVSSPVEG-----KDEGRLRLRLRYA 132
 DB 1 MSPLGFTVNLVVRWQEQRLPWPDTLALFSAIPNSNGVNNETGLMRNIMRYM 60
 QY 133 ILQVLILRSISVYKRPFTLHLVLGPMTHGEHKQLKLG--LPNHTFWVPWVFAN 190
 DB 61 VLAYVITLQRLSRVRRPPTTQHLVDAGLHSEMKIPALNQKSPMSKYWPLWATN 120
 QY 191 LSKAVLGRIRDTVLQSLMNEVCTLRTOCGOLYAYDWISIPLVYTVQVTVAYGFFLA 250
 DB 121 IINARKDGLIASDHIVQTILVELSDIRRLGLGLIGYDTCVPLVYTVQVTVLYVYFA 180
 QY 251 CLIGROFLNPKDYPGHE--MDLVVPVFTILQFLFYMGWLKVAEQLNPFGEDDDDPFTNW 309
 DB 181 ALLGRQMLNVLDRSGEDPDLFFLFTVLQFLFYMGWLKVAEVLNPFGEDDDDDLNKN 240
 QY 310 IIDRLNQLVSLSDVGMHQNLPMPERDMYNEAAPQP--PYTAASARRRHSFMGSTENISL 368
 DB 241 LIDRHIKAAYMIVDENHEEHPDLLRQYWCVVPKDLPYTVASEHYRKDEPKSAEKYV 300
 QY 369 KKED 372
 DB 301 KKED 304
 RESULT 9
 AAB70085
 ID AAB70085 standard; Protein; 314 AA.
 XX
 AC AAB70085;
 XX
 DT 14-MAY-2001 (first entry)
 XX
 DE Human secreted protein-related polypeptide #6.
 XX
 KW Human; secreted protein; immunomodulatory; antisclerotic;
 KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
 KW antialzheimers; antiparkinsonian; antimicrobial; vulnery; gene therapy;
 KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
 KW neurological; infection.
 OS
 XX Homo sapiens.

PN WO200112776-A2.
 XX
 PD 22-FEB-2001.
 XX
 PF 15-AUG-2000; 2000WO-US22350.
 XX
 PR 16-AUG-1999; 99US-0148759.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;
 DR WPI; 2001-244245/25.
 XX
 PT Nucleic acids encoding 18 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX
 PS Disclosure; Page 377-378; 380pp; English.
 XX
 CC The present sequence is given in a specification relating to nucleic
 CC acid molecules encoding 18 novel human secreted proteins. The nucleic
 CC acids and proteins may be used in the prevention, diagnosis and
 CC treatment of diseases including immune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
 CC Chagas's cardiomyopathy and coronary arteriosclerosis), angiogenic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. The
 CC nucleic acid molecules may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples. The polypeptides may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators of
 CC protein expression and activity.
 XX
 SQ Sequence 314 AA;
 Query Match 21.8%; Score 641; DB 22; Length 314;
 Best Local Similarity 53.8%; Pred. No. 1.3e-57;
 Matches 129; Conservative 31; Mismatches 62; Indels 18; Gaps 4;
 QY 179 NTFWVPWVNFANLSMKAYLGRIRDTVLQSLMNEVCTLRTOCGOLYAYDWISIPLVYTO 238
 DB 20 NKYVPCVWFTNLAQARRDGRIRDDIALCLLELNKTRAKCSMLFHDWISIPLVYTO 79
 QY 239 VVTVAVYSPFLACLIGRQFLNPNKD-----YPGHE-----MDLVVPVFTILQFLF 283
 DB 80 VVTIAVYSPFALSIVGRQFVEPEAGAAPKQLKPGQEPAPALGDFMVPVLTLLQFPF 139
 QY 284 YMGWLKVAEQLNPFGEDDDDPFTNWIIDRLNQLVSLSDVGMHQNLPMPERDMYNEAAP 343
 DB 140 YAGWLKVAEQIINPFGEDDDDPFTNQLIDRLNQLVSLSDVGMHQNLPMPERDMYNEAAP 199
 QY 344 QPPYTT--AASARRRHSFMGSTENISLKKEDSLWSKEADTKKESGYST--IGCFGLG 400
 DB 200 QPPYTVATAESLSPFLGSTFNLRMSDDPEQSLQVEASPGSGRPAPAPQTPLLGRFLGV 259
 RESULT 10
 AAB70067
 ID AAB70067 standard; Protein; 251 AA.
 XX
 AC AAB70067;
 XX
 DT 14-MAY-2001 (first entry)
 XX
 DE Human secreted protein #6.
 XX
 KW Human; secreted protein; immunomodulatory; antisclerotic;
 KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;

KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
 KW antialzheimers; antiparkinsonian; antimicrobial; vulnerrary; gene therapy;
 KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
 KW neurological; infection.

OS Homo sapiens.

PN WO200112776-A2.

XX 22-FEB-2001.

XX 15-AUG-2000; 2000WO-US223350.

XX 16-AUG-1999; 99US-0148759.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;

XX WPI: 2001-244245/25.

DR N-PSDB; AAF76848.

XX Nucleic acids encoding 18 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -

PS Claim 11; Page 362-363; 380pp; English.

XX The present sequence is one of 18 novel human secreted proteins. The
 CC nucleic acids encoding the proteins and the proteins themselves may be
 CC used in the prevention, diagnosis and treatment of diseases including
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
 CC and human immunodeficiency virus (HIV) infections), hyperproliferative
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
 CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC infectious diseases and/or for promoting wound healing, regeneration
 CC and/or chemotaxis. The nucleic acid molecules may also be used as DNA
 CC probes in diagnostic assays to detect and quantitate the presence of
 CC similar nucleic acid sequences in samples. The polypeptides may also be
 CC used as antigens in the production of antibodies and in assays to
 CC identify modulators of protein expression and activity.

XX Sequence 251 AA;

Query Match 18.1%; Score 534; DB 22; Length 251;

Best Local Similarity 55.9%; Pred. No. 1.2e-46;

Matches 109; Conservative 24; Mismatches 44; Indels 18; Gaps 4;

QY 224 LYAVDWISIPLVYTVQVTVAVYSPFALSLVGRQFVEPEAGAAKPKQLKPGQEPAPALGD 61

Db 2 LFHYDWISIPLVYTVQVTVAVYSPFALSLVGRQFVEPEAGAAKPKQLKPGQEPAPALGD 61

QY 269 MDLVVVPVFTILOFLPYMGWLKVAEQLINPFGEEDDDFETNWIIDRNQLVSLSDVMGHON 328

Db 62 PDMVPLTLLQFFFYAGWLKVAEQIINPFGEEDDDFETNQLIDRNQLVSLSDVMYQON 121

QY 329 LPPMERDMYNEAAPQPPYT-AAASRRSRHSGFMSTFNISLKKEDELEWSKEEADTKKE 387

Db 122 LPPAEKQYWDQDQPPYTVATAESLRPFLSGTFLNLRMSDDPEQSLQVFAEFGSGRP 181

QY 388 SGYSST--IGCFGLG 400

Db 182 APAAQTELLGRFLGV 196

RESULT 11

ABG55506

ID ABG55506 standard; Protein; 251 AA.

XX

AC

XX

DT

XX

DE

XX

KW

KW

KW

KW

KW

KW

KW

KW

KW

OS

OS

XX

XX

PN

XX

PD

XX

XX

PF

XX

PR

XX

PR

XX

XX

PA

XX

XX

PI

XX

XX

DR

XX

XX

PT

XX

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

ABG55506;

27-AUG-2002 (first entry)

Human albumin fusion protein #2181.

Albumin fusion protein; therapeutic protein X; human albumin; HA;
 human serum albumin; HSA; cancer; reproductive disorder;
 digestive disorder; immune disorder; endocrine disorder;
 haematopoietic disorder; neural disorder; connective disorder;
 cytostatic; antifertility; antiinflammatory; antiulcer;
 immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 osteopathic; antiarthritic.

Homo sapiens.

Synthetic.

WO200177137-A1.

18-OCT-2001.

12-APR-2001; 2001WO-US11988.

12-APR-2000; 2000US-229358P.

25-APR-2000; 2000US-199384P.

21-DEC-2000; 2000US-256931P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Haseltine WA;

WPI: 2002-010886/01.

New fusion protein for treating disease e.g. diabetes comprises an
 albumin fused to a therapeutic protein -

Claim 1; Page 2075; 2102pp; English.

The present invention relates to albumin fusion proteins comprising a
 therapeutic protein X and human albumin (HA, also known as human serum
 albumin, HSA). The proteins are useful for treating a disease or
 disorder that may be modulated by therapeutic protein X. The albumin
 extends the shelf-life of protein X, and may increase its biological
 in vitro/in vivo activity. The protein is useful for treating and
 diagnosing disorders such as cancer, reproductive disorders, digestive
 disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 (e.g. diabetes), haematopoietic disorders, neural disorders
 (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 encephalomyelitis, meningitis, schizophrenia) and connective disorders
 (e.g. osteoporosis, arthritis). ABG53326-ABG55518 represent albumin
 fusion proteins of the invention.

Sequence 251 AA;

Query Match

Best Local Similarity 18.1%; Score 534; DB 23; Length 251;

Matches 109; Conservative 24; Mismatches 44; Indels 18; Gaps 4;

QY 224 LYAVDWISIPLVYTVQVTVAVYSPFALSLVGRQFVEPEAGAAKPKQLKPGQEPAPALGD 61

Db 2 LFHYDWISIPLVYTVQVTVAVYSPFALSLVGRQFVEPEAGAAKPKQLKPGQEPAPALGD 61

QY 269 MDLVVVPVFTILOFLPYMGWLKVAEQLINPFGEEDDDFETNWIIDRNQLVSLSDVMGHON 328

Db 62 PDMVPLTLLQFFFYAGWLKVAEQIINPFGEEDDDFETNQLIDRNQLVSLSDVMYQON 121

QY 329 LPPMERDMYNEAAPQPPYT-AAASRRSRHSGFMSTFNISLKKEDELEWSKEEADTKKE 387

Db 122 LPPAEKQYWDQDQPPYTVATAESLRPFLSGTFLNLRMSDDPEQSLQVFAEFGSGRP 181

QY 388 SGYSST--IGCFGLG 400

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-541565/60.
DR N-PSDB; ABA11094.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Claim 11; SEQ ID NO 3425; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (AB14678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 101 AA;

Query Match 12.7%; Score 374; DB 22; Length 101;
Best Local Similarity 74.5%; Pred. No. 1.3e-30;
Matches 70; Conservative 9; Mismatches 15; Indels 0; Gaps 0;
QY 291 AEQLINPEGDDDDFFETNWIIDRLNQLVSLSDVGHQNLPMERDMYNEARPPPTAA 350
DB 1 AEQLINPEGDDDDFFETNWIIDRLNQLVSLSDVGHQNLPMERDMYNEARPPPTAA 60
QY 351 SARRRRHSFMGSTFNISLKEDLELWSKEADTD 384


```

PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0233423.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 03-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(PA) (HUMA-) HUMAN GENOME SCI INC.
(PI) Rosen CA, Barash SC, Ruben SM;
(XX) WPI; 2001-465567/50.
(XX) N-PSDB; AAS39532.

Isolated polypeptide for treating, preventing and/or prognosing
disorders related to the colon including colon cancers and also for
testing and detection e.g. diagnosis -

Claim 1; SEQ ID No 429; 562pp; English.

The present invention relates to the isolation of novel human colon
associated polypeptides, and the cDNA (AAS39348-AAS39581) and genomic
sequences encoding for them. The sequences of the invention are useful
in the diagnosis, treatment, prevention and/or prognosis of disorders
of the colon including colon cancer, congenital abnormalities
(e.g. atresia and stenosis), bacterial and viral infections,
inflammatory bowel disease (IBD), neoplastic cell disorders,
colonic inflammation, adenomas, intestinal inflammatory disorders, colitis,
(e.g. polyps and adenomas), diarrhoea and dysentery, malabsorption syndromes
(e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
The polynucleotide sequences of the invention can also be used in gene
therapy. AAU22468-AAU22701 represent the novel human colon associated
polypeptides of the invention.

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Query Match 7.9%; Score 232; DB 22; Length 99;
Best Local Similarity 55.4%; Pred. No. 7.1e-16;
Matches 46; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 1 MTITYTNKVNARLGSFSSLLCWRGSIYKLLYGFVFIYLSIRGLYVMSDDQL 60
Db 13 MTVSYTLKVAERRGSGFLLLRGSIYKLLYKFLFGALYAVLSITYRLLTQGRY 72
QY 61 LFEKALYCDSYIQIIPISFVLG 83
Db 73 VYAOVARYCNRSADLIPLSFVLG 95

```

Search completed: July 10, 2003, 12:24:13
Job time : 35.021 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:34 ; Search time 12.6264 Seconds
(without alignments)
1283.984 Million cell updates/sec

Title: US-09-622-964-29

Perfect score: 2944

Sequence: 1 MTITYTNKANVARLGSFSSL.....EHASSPYRDEAGTKPVLVE 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCITUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96.5	3.3	839	4	US-09-197-636-2
2	96.5	3.3	839	4	US-09-197-636-4
3	96.5	3.3	839	4	US-09-197-636-8
4	96.5	3.3	839	4	US-09-235-451-34
5	96.5	3.3	839	4	US-09-533-220A-2
6	91.5	3.1	993	1	US-08-468-557-2
7	89.5	3.0	838	4	US-09-235-451-2
8	89.5	3.0	838	4	US-09-132-316-3
9	89	3.0	1099	4	US-09-442-100-2
10	88.5	3.0	997	1	US-08-233-540-1
11	88.5	3.0	997	1	US-08-428-949A-1
12	88.5	3.0	997	1	US-08-428-948A-1
13	88.5	3.0	997	2	US-08-428-946-1
14	88.5	3.0	997	5	PCT-US95-04656-1
15	88.5	3.0	998	1	US-08-233-008A-6
16	88.5	3.0	1021	2	US-08-233-008A-2
17	88	3.0	688	2	US-09-016-000-2
18	87	3.0	681	2	US-08-272-255-9
19	87	3.0	681	2	US-08-272-255-6
20	87	3.0	681	5	PCT-US95-08565-6
21	87	3.0	681	5	PCT-US95-08565-9
22	86.5	2.9	2304	4	US-09-324-867-4
23	86	2.9	1956	4	US-08-843-417-10
24	86	2.9	2319	1	US-08-213-133A-8
25	86	2.9	2319	1	US-08-474-503-6
26	86	2.9	2319	2	US-08-670-707A-6
27	86	2.9	2319	4	US-09-037-601-6

28 86 2.9 2319 4 US-09-315-179-6 Sequence 6, Appli
29 86 2.9 2319 4 US-09-523-656-28 Sequence 28, Appli
30 86 2.9 2319 5 PCT-US94-13200-6 Sequence 6, Appli
31 85.5 2.9 459 4 US-09-097-889-22 Sequence 22, Appli
32 85.5 2.9 684 4 US-09-134-078-62 Sequence 62, Appli
33 84.5 2.9 968 4 US-09-302-812-6 Sequence 6, Appli
34 84.5 2.9 968 4 US-09-511-477-6 Sequence 6, Appli
35 84.5 2.9 968 4 US-09-511-507-6 Sequence 6, Appli
36 84.5 2.9 1144 1 US-08-147-812-5 Sequence 5, Appli
37 84.5 2.9 1144 2 US-08-319-866-12 Sequence 12, Appli
38 84.5 2.9 1144 4 US-09-123-708-2 Sequence 2, Appli
39 84.5 2.9 1144 4 US-09-123-624-2 Sequence 2, Appli
40 84 2.9 681 3 US-08-984-268-5 Sequence 5, Appli
41 84 2.9 1984 3 US-08-836-325-10 Sequence 10, Appli
42 83.5 2.8 237 2 US-08-760-745-5 Sequence 5, Appli
43 83.5 2.8 554 4 US-09-134-001C-5109 Sequence 5109, Ap
44 83 2.8 626 2 US-08-617-801A-4 Sequence 4, Appli
45 83 2.8 857 1 US-07-717-331F-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-197-636-2
; Sequence 2, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-197-636-2

```
Query Match      3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

QY 26 GSIYKLYGFLVFIFLYYSIRGLYRMVLSDDQQLFEKALYCDSYIQLIPISFVLGFY 85
DB 477 GEILSVLGG-----VYFFRGIQYEL-----QRRPSMKTFLFVDSYSEM--LFFLOSFL 522
QY 86 VTLVSRWMOYEN-----LPWDRMIQVSSFVEGKDEEGRLLRRLTLRYAILG 135
DB 523 MLATVVLVYFSLHKEYVASMVFSALGWTNMLY-----YTRGFQMG-----IYAVMI 569
QY 136 QVILIRISISVYKRPFTLHLVLGFM-----HGEHKQLQKLGILPHNTFWVPWVF 188
DB 570 EKMLRLDLC-----RPMFVIVFLFGSTAVVTLIEDGKNDLSPSESHR----- 615
QY 236 YQVTVVAVYSFPLACLIQROFLPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLAYVILTITLLNMLIALMGETVKNIAQESKNIKLQRAITILDTEKSLKCMKRAFR 722
QY 285 MGWLKVAEQLINPGEDDDF-----ETNWI-----IDRLQVSLLS 321
DB 723 SG--KLLQVGYTPDGKDDYRCFRVDEVNWTNTNVTNVIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMDERDMYNEAAPQPPYTAASARSRHSFMGST-----FNISLKKEDELEL 375
DB 781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLRPEDA EV 829
QY 376 WSKKEADTK 385
DB 830 FKSPAASGEK 839

RESULT 2
US-09-197-636-4
; Sequence 4, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
```

```
NAME: Prestia, Paul P
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-30075
TELECOMMUNICATION INFORMATION:
TELEPHONE: 601-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 839 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-197-636-4

Query Match      3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

QY 26 GSIYKLYGFLVFIFLYYSIRGLYRMVLSDDQQLFEKALYCDSYIQLIPISFVLGFY 85
DB 477 GEILSVLGG-----VYFFRGIQYEL-----QRRPSMKTFLFVDSYSEM--LFFLOSFL 522
QY 86 VTLVSRWMOYEN-----LPWDRMIQVSSFVEGKDEEGRLLRRLTLRYAILG 135
DB 523 MLATVVLVYFSLHKEYVASMVFSALGWTNMLY-----YTRGFQMG-----IYAVMI 569
QY 136 QVILIRISISVYKRPFTLHLVLGFM-----HGEHKQLQKLGILPHNTFWVPWVF 188
DB 570 EKMLRLDLC-----RPMFVIVFLFGSTAVVTLIEDGKNDLSPSESHR----- 615
QY 189 ANLSMKAYLGRIRDTVLLQSLMNEVCTLRQCQOLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP---DSSYN---SLYSTCLELKFKTIGMGDLFTENYDFKAVFII 662
QY 236 YQVTVVAVYSFPLACLIQROFLPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLAYVILTITLLNMLIALMGETVKNIAQESKNIKLQRAITILDTEKSLKCMKRAFR 722
QY 285 MGWLKVAEQLINPGEDDDF-----ETNWI-----IDRLQVSLLS 321
DB 723 SG--KLLQVGYTPDGKDDYRCFRVDEVNWTNTNVTNVIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMDERDMYNEAAPQPPYTAASARSRHSFMGST-----FNISLKKEDELEL 375
DB 781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLRPEDA EV 829
QY 376 WSKKEADTK 385
DB 830 FKSPAASGEK 839

RESULT 3
US-09-197-636-8
; Sequence 8, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

us-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

```

; APPLICANT: Caterina, Michael J.
; APPLICANT: Brake, Anthony J.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 9076/084CIP
; CURRENT APPLICATION NUMBER: US/09/235,451
; CURRENT FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/072,151
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 08/915,461
; PRIOR FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-235-451-34

Query Match      3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74; Indels 137; Gaps 19;
Matches 85; Conservative 54; Mismatches 54;

QY      26 GSIYKLLYGEFLVFIFLYYSIRGLYRMYLSSDQQLLEKALYCDSYIQLIPISFVLGFI 85
      477 GEILSVLGG-----VYFFFRGQIYFL-----QRRPSMKTFLFVDSYSEM--LFFLQSLF 522
      86 VTLVSRMWSQYEN-----LPPDRMLQVSSFVEGKDEBGLLRRLRLIRYAILG 135
      523 MLATVLYFSLHKEYVYASVMPFSLALGWTNMLY-----YTRGFQOMG-----IYAVMI 569
      136 QVILRSISTSVKRPFTLHLVLAGEMT-----HGEHKQLQKLGPHNPTFWVWVWF 188
      570 EKMLRLDLC-----RPMFYVVFVLPFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
      189 ANLSMKAYLGGIRDTVLLQSLMNEVCTLRTOCGQLYA-----YDWISIPLV 235
      616 -----WRGPACRPP-----DSSYN---SLYSTCLELFTKFTIGMGDLFTENYDFKAVFII 662
      236 YQVTVVAVYSFPLACLIHQFLNPKNDYPGHE-----MDLVVPVFTILOFLFY 284
      663 LLLAYVILTYILLNMLIALMGETVKNKIAQESKNWKLQRAITILDTEKSFCLKMKRAFR 722
      285 MGWLKVAEQLINPGEDDDF-----ETNMI-----IDRLQVSLLS 321
      723 SG--KLQVGYTPDGKDYRCFVDEVNMTTNTNVGIINEDPGNCEGVKRTLSFSLRS 780
      322 --VDGMHQNLPFMRDMYNEAAPQPPYTAASARRSRHSFMGST-----FNISLKEDLEL 375
      781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPDAEV 829
      376 WSKKEADTDK 385
      830 FKSPAASGEK 839

RESULT 5
US-09-533-220A-2
; Sequence 2, Application US/09533220A
; Patent No. 6406908
; GENERAL INFORMATION:
; APPLICANT: McIntyre, Peter
; APPLICANT: James, Iain Fraser
; TITLE OF INVENTION: Human Vanilloid Receptor
; FILE REFERENCE: 4-30875A
; CURRENT APPLICATION NUMBER: US/09/533,220A
; CURRENT FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: UNITED KINGDOM 9907097.1
; PRIOR FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 1.30
; SEQ ID NO 2

```

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-197-636-8

Query Match      3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74; Indels 137; Gaps 19;
Matches 85; Conservative 54; Mismatches 54;

QY      26 GSIYKLLYGEFLVFIFLYYSIRGLYRMYLSSDQQLLEKALYCDSYIQLIPISFVLGFI 85
      477 GEILSVLGG-----VYFFFRGQIYFL-----QRRPSMKTFLFVDSYSEM--LFFLQSLF 522
      86 VTLVSRMWSQYEN-----LPPDRMLQVSSFVEGKDEBGLLRRLRLIRYAILG 135
      523 MLATVLYFSLHKEYVYASVMPFSLALGWTNMLY-----YTRGFQOMG-----IYAVMI 569
      136 QVILRSISTSVKRPFTLHLVLAGEMT-----HGEHKQLQKLGPHNPTFWVWVWF 188
      570 EKMLRLDLC-----RPMFYVVFVLPFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
      189 ANLSMKAYLGGIRDTVLLQSLMNEVCTLRTOCGQLYA-----YDWISIPLV 235
      616 -----WRGPACRPP-----DSSYN---SLYSTCLELFTKFTIGMGDLFTENYDFKAVFII 662
      236 YQVTVVAVYSFPLACLIHQFLNPKNDYPGHE-----MDLVVPVFTILOFLFY 284
      663 LLLAYVILTYILLNMLIALMGETVKNKIAQESKNWKLQRAITILDTEKSFCLKMKRAFR 722
      285 MGWLKVAEQLINPGEDDDF-----ETNMI-----IDRLQVSLLS 321
      723 SG--KLQVGYTPDGKDYRCFVDEVNMTTNTNVGIINEDPGNCEGVKRTLSFSLRS 780
      322 --VDGMHQNLPFMRDMYNEAAPQPPYTAASARRSRHSFMGST-----FNISLKEDLEL 375
      781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPDAEV 829
      376 WSKKEADTDK 385
      830 FKSPAASGEK 839

RESULT 4
US-09-235-451-34
; Sequence 34, Application US/09235451
; GENERAL INFORMATION:
; APPLICANT: Julius, David J.

```

us-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

```

NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI94-13A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 993 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-557-2

Query Match 3.1%; Score 91.5; DB 1; Length 993;
Best Local Similarity 19.6%; Pred. No. 3.2;
Matches 83; Conservative 53; Mismatches 134; Indels 153; Gaps 20;

QY 94 WSOYENLPWDRMIQVSSFVEGKDEGRLLRRLIRYAILGQVLIRLSISVYKRPPT 153
DB 588 WSYV-LFGPDQ---RADLYEGKQQLGWFQSSLL-----TSVAARKKA 627
QY 154 LHLVLVAGMTGHE--HKQLQKLG-----LPHNTFVWPVWVFA--NL 191
DB 628 PYKTVIVGFTLGEKGMKSLGNVHPDVVNGGQDQSKPEPPYGVADVLRW--WVADSNV 686
QY 192 SMKAYLG-----GRIDTV-----LLQS 209
DB 687 FTEVAIGPSVLNAARDITSLKRLTLRFLGNVADFNPTDTSIPVNDMYVIDQYMLHLQD 746
QY 210 LKNEVCTLRQCGQYAYDWISIPLVYVTVVAVYFFLACLGROFLNPKNDYPGHE- 268
DB 747 LANKITELYKQ-----YDFGVVVELLRTFTYRELSNFFSIIKDRLYCEKENDPKRSC 800
QY 269 -----MDLVVPVFT-ILQFLFYMGWLKVAEOLIN--PF-GEDDDDPFETWIDRLQV 317
DB 801 QTALVEILDVIVRSFAPILPHL-----AEVQHPIYIPEKPSVFTGTWISTSI-- 850
QY 318 SLLSVGDGHQNLPPMERDMYVNEAAPQPPYTAASARSRRHSPMGS-----TF 364
DB 851 -----WKK--PGLBEAVESACARMSDFLSIPGKNAAEYKVIIV 888
QY 365 NISLKEDLE-LMSKEADTKKESGYSSTIGCFLQPKNYHLPLKDLTKLCKSNPL 423
DB 889 EPGLLFEIEMLOSEETSSQLNELMASESTLLAQEPREMTADVIELKGRFLIN---- 944
QY 424 LEG 426
DB 945 LEG 947

RESULT 7
US-09-235-451-2
Sequence 2, Application US/09235451
GENERAL INFORMATION:
APPLICANT: Julius, David J.
APPLICANT: Caterina, Michael J.
APPLICANT: Brake, Anthony J.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED
POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 9076/084CIP
CURRENT APPLICATION NUMBER: US/09/235,451
CURRENT FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/072,151
PRIOR FILING DATE: 1998-01-22
PRIOR APPLICATION NUMBER: 08/915,461
PRIOR FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 838
TYPE: PRT

```

us-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

```

ORGANISM: R. rattus
US-09-235-451-2
Query Match
Best Local Similarity 18.5%; Pred. No. 3.9; Indels 165; Gaps 22;
Matches 87; Conservative 61; Mismatches 157; Indels 165; Gaps 22;

486 VYFFPRGQVFLQRRPSLKSFLVDSYSEI--LFFVQSLFMLVSVVLYFSQKEYVASNVF 543
100 ---LPWPDRLMIQVSSPVGKDBEGRLRLRLIRVAILGQVLILRSISTSVYKRPPTLHH 156
544 SLAMGWTNMLY-----YTRGFQMG-----IYAVMIKMLRDLC-----RPMFVYL 585
157 LVLAGFMT-----HGEHKQKQKGLPHNTFWPWWFANLSMKAYLGGRIKRTDVLLOS 209
586 VFLGFGSTAVVTLIEDGKNSLPMESTPH-----KCRGSACKPG 624
210 LMNEVCTLRTOCGOLYA-----YDWISIPLVYVTVVTVVAVVSFFLACLIQRQ 256
625 ---NSYNSLYSTCLFELFKFTMGDLFTENYDKAVFIILLAYVILTYILLINMLIALM 682
257 FLNPNKQYPGHE-----MDLVVPVFTILQFLFYMGWLKVAEOLINPFGEDDDDF 305
683 GETVNKIAQESKNWIKLQRAITILDTEKSLKMKRKAFRSG--KLQVGFPTDGKDDYRW 740
306 -----ETWNI-----IDRNQVSLLS--VDGMHQNLPMPERDMYNE 340
741 CFPVDEVNVTWNTVNGIINEDPGNCGVGKRTLSFSLRSGRVSGRN-----WKN 789
341 AAPQPPYTAASARRSRHSGFMSTFNI-----SLKKEDELEMSKEEADTK 385
790 FALVPLLRDASTRD--RHATQOEVLKHVTGSLKPEDAEVFKDSNVPGEK 838

RESULT 9
US-09-442-100-2
; Sequence 2, Application US/09442100
; Patent No. 6359193
; GENERAL INFORMATION:
; APPLICANT: Xu, Tian
; APPLICANT: Tao, Wufan
; APPLICANT: Wang, Weiyl
; APPLICANT: Zhang, Sheng
; APPLICANT: Yu, Wan
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
; TITLE OF INVENTION: GENES AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442.100
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/411,111
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6523-003
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1099 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
;

24 WRGSIKLYGBFLVP-----IF-----LYSIRGL--YRM-----VLSSDQ 59
426 WDRFVKRIFFNFYVCLYMIIFTAAAYRVPVEGLPPYKLNKNTVGDYFRVTGEILSVSGG 485
60 LLF-----EKALYCDSYIQLIPISFVLGFTVTVLTVSRMWSQYEN----- 99
486 VYFFPRGQVFLQRRPSLKSFLVDSYSEI--LFFVQSLFMLVSVVLYFSQKEYVASNVF 543
100 ---LPWPDRLMIQVSSPVGKDBEGRLRLRLIRVAILGQVLILRSISTSVYKRPPTLHH 156
544 SLAMGWTNMLY-----YTRGFQMG-----IYAVMIKMLRDLC-----RPMFVYL 585
157 LVLAGFMT-----HGEHKQKQKGLPHNTFWPWWFANLSMKAYLGGRIKRTDVLLOS 209
586 VFLGFGSTAVVTLIEDGKNSLPMESTPH-----KCRGSACKPG 624
210 LMNEVCTLRTOCGOLYA-----YDWISIPLVYVTVVTVVAVVSFFLACLIQRQ 256
625 ---NSYNSLYSTCLFELFKFTMGDLFTENYDKAVFIILLAYVILTYILLINMLIALM 682
257 FLNPNKQYPGHE-----MDLVVPVFTILQFLFYMGWLKVAEOLINPFGEDDDDF 305
683 GETVNKIAQESKNWIKLQRAITILDTEKSLKMKRKAFRSG--KLQVGFPTDGKDDYRW 740
306 -----ETWNI-----IDRNQVSLLS--VDGMHQNLPMPERDMYNE 340
741 CFPVDEVNVTWNTVNGIINEDPGNCGVGKRTLSFSLRSGRVSGRN-----WKN 789
341 AAPQPPYTAASARRSRHSGFMSTFNI-----SLKKEDELEMSKEEADTK 385
790 FALVPLLRDASTRD--RHATQOEVLKHVTGSLKPEDAEVFKDSNVPGEK 838

RESULT 8
US-09-132-316-3
; Sequence 3, Application US/09132316B
; Patent No. 6444440
; GENERAL INFORMATION:
; APPLICANT: Young, Steven M.
; APPLICANT: Ruben, Vanilloid Receptor-2
; TITLE OF INVENTION: Vanilloid Receptor-2
; FILE REFERENCE: 1488.1110000
; CURRENT APPLICATION NUMBER: US/09/132.316B
; EARLIER FILING DATE: 1998-08-11
; EARLIER FILING DATE: 1997-03-07
; EARLIER FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-132-316-3
Query Match
Best Local Similarity 18.5%; Pred. No. 3.9; Indels 165; Gaps 22;
Matches 87; Conservative 61; Mismatches 157; Indels 165; Gaps 22;

24 WRGSIKLYGBFLVP-----IF-----LYSIRGL--YRM-----VLSSDQ 59
426 WDRFVKRIFFNFYVCLYMIIFTAAAYRVPVEGLPPYKLNKNTVGDYFRVTGEILSVSGG 485
60 LLF-----EKALYCDSYIQLIPISFVLGFTVTVLTVSRMWSQYEN----- 99
486 VYFFPRGQVFLQRRPSLKSFLVDSYSEI--LFFVQSLFMLVSVVLYFSQKEYVASNVF 543
100 ---LPWPDRLMIQVSSPVGKDBEGRLRLRLIRVAILGQVLILRSISTSVYKRPPTLHH 156
544 SLAMGWTNMLY-----YTRGFQMG-----IYAVMIKMLRDLC-----RPMFVYL 585
157 LVLAGFMT-----HGEHKQKQKGLPHNTFWPWWFANLSMKAYLGGRIKRTDVLLOS 209
586 VFLGFGSTAVVTLIEDGKNSLPMESTPH-----KCRGSACKPG 624
210 LMNEVCTLRTOCGOLYA-----YDWISIPLVYVTVVTVVAVVSFFLACLIQRQ 256
625 ---NSYNSLYSTCLFELFKFTMGDLFTENYDKAVFIILLAYVILTYILLINMLIALM 682
257 FLNPNKQYPGHE-----MDLVVPVFTILQFLFYMGWLKVAEOLINPFGEDDDDF 305
683 GETVNKIAQESKNWIKLQRAITILDTEKSLKMKRKAFRSG--KLQVGFPTDGKDDYRW 740
306 -----ETWNI-----IDRNQVSLLS--VDGMHQNLPMPERDMYNE 340
741 CFPVDEVNVTWNTVNGIINEDPGNCGVGKRTLSFSLRSGRVSGRN-----WKN 789
341 AAPQPPYTAASARRSRHSGFMSTFNI-----SLKKEDELEMSKEEADTK 385
790 FALVPLLRDASTRD--RHATQOEVLKHVTGSLKPEDAEVFKDSNVPGEK 838

Query Match
Best Local Similarity 18.5%; Pred. No. 3.9; Indels 165; Gaps 22;
Matches 87; Conservative 61; Mismatches 157; Indels 165; Gaps 22;

24 WRGSIKLYGBFLVP-----IF-----LYSIRGL--YRM-----VLSSDQ 59
426 WDRFVKRIFFNFYVCLYMIIFTAAAYRVPVEGLPPYKLNKNTVGDYFRVTGEILSVSGG 485
60 LLF-----EKALYCDSYIQLIPISFVLGFTVTVLTVSRMWSQYEN----- 99
486 VYFFPRGQVFLQRRPSLKSFLVDSYSEI--LFFVQSLFMLVSVVLYFSQKEYVASNVF 543
100 ---LPWPDRLMIQVSSPVGKDBEGRLRLRLIRVAILGQVLILRSISTSVYKRPPTLHH 156
544 SLAMGWTNMLY-----YTRGFQMG-----IYAVMIKMLRDLC-----RPMFVYL 585
157 LVLAGFMT-----HGEHKQKQKGLPHNTFWPWWFANLSMKAYLGGRIKRTDVLLOS 209
586 VFLGFGSTAVVTLIEDGKNSLPMESTPH-----KCRGSACKPG 624
210 LMNEVCTLRTOCGOLYA-----YDWISIPLVYVTVVTVVAVVSFFLACLIQRQ 256
625 ---NSYNSLYSTCLFELFKFTMGDLFTENYDKAVFIILLAYVILTYILLINMLIALM 682
257 FLNPNKQYPGHE-----MDLVVPVFTILQFLFYMGWLKVAEOLINPFGEDDDDF 305
683 GETVNKIAQESKNWIKLQRAITILDTEKSLKMKRKAFRSG--KLQVGFPTDGKDDYRW 740
306 -----ETWNI-----IDRNQVSLLS--VDGMHQNLPMPERDMYNE 340
741 CFPVDEVNVTWNTVNGIINEDPGNCGVGKRTLSFSLRSGRVSGRN-----WKN 789
341 AAPQPPYTAASARRSRHSGFMSTFNI-----SLKKEDELEMSKEEADTK 385
790 FALVPLLRDASTRD--RHATQOEVLKHVTGSLKPEDAEVFKDSNVPGEK 838

```

Fri Jul 11 11:45:18 2003

US-09-442-100-2

```

Query Match      3.0%; Score 89; DB 4; Length 1099;
Best Local Similarity 20.3%; Pred. No. 6.9;
Matches 96; Conservative 64; Mismatches 162; Indels 152; Gaps 25;

101 PMPDRLMIQVSVFVGKDEBGRLLRRLIRYAILGOVLLRSISTSVYKRPFTLH-HLVL 159
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
619 PIPERKIS-----KXKEEERKEFR-----IRQYSPQAFKPFMECHTENV 659
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
160 AGF--MTHGEH--KOLQKLGHPHNTFWVPWFWFANLSMKAYLGGRIQRTDVLLO----- 208
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
660 KSYRQRTYRNQLEKEMHKVGLPDQT-----QIEMRMLNQKESNYIRLGRKADK 710
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 -----SLANVECTLRTOCGOLYAYDWT--SIPLVYTVVTVAVYSFFL 249
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
711 SMFVKLPGVGAFFGVTLVSKIDT-----SNHLYAMKTLRKADVLKRNQVAHVKAERDIL 766
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
250 A-----CLIGROFLNPKO-----YPCHE-MDLVWVPVFTILOPL--FYMGWLKVAEQ 293
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
767 AEDNNWVLYYSFQKONLYFMDYIPGGDLMSLLIKLIGIFEEELARFYAETCAVD 826
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
294 LINPFGEDDDDFE--TNWIIRNLOVSLSV-----DCMHQNLPPMERD 335
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
827 SVHKMGPIHRDIKPDNIIIDRDGHILKLTDFGLCTGFRWTHNSKYQENGNHSRQDSMEPW 886
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 MYWNEAARPPYTAASARSR-----HSFPMGSTFNISLKEDLELMSKEEADTKKES 388
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
887 BEYSENGPKP--TVLERMRDHORVLAHSLVGTENVIA--PEVLE-----RS 930
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
389 GYSS-----TIGC-----FLGLOP-----KNVHLPLDKLTKLLCSKN 421
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
931 GYTQLCDYSGVGVILYELMVGOPPLANSPLETOOKVINWEKTHIP---PQAEISREAT 987
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
422 PILEGOCQANQKQKQVWKFGLDFLKCVPRFKRGSHGCPQAPS--SHPTQOS 474
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
988 DLIRLCASADKRLGKSVDEVKSHDFPKGID--FADMRKQKAPYIPEIKHPTDTS 1040
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10

```

US-08-232-540-1
; Sequence 1, Application US/08232540
; Patent No. 5498536
; GENERAL INFORMATION:
; APPLICANT: Khandke, Kiran M.
; TITLE OF INVENTION: A No. 5498536el Protein Designated
; TITLE OF INVENTION: Chondroitinase II and its Use With A Protein Designated
; TITLE OF INVENTION: Chondroitinase I to Achieve Complete Vitreal Disinsertion
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,540
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 32,390
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3244
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ.ID NO: 1:

```

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 997 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-232-540-1

```

Query Match 3.0%; Score 88.5; DB 1; Length 997;

```

Best Local Similarity 19.0%; Pred. No. 6.6;
Matches 87; Conservative

```

```

55 SSDQQLPEKALYCDSYQIQLIPISFVLGYVTVLVVSRWMSQYENLPWDRMLMIQVS-SF 113
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
300 SODKQLF-----DNYV-----ILGNYTT-----LMFNISRAY 326
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
114 VBGKDEBGRLLRRLIRYAILGOVLLRSISTSVYKRPFTLHHLVLGFMTHGHEKQLOK 173
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 VLEKDPQTK-----AQLKQMYLLMT-----KHLDDQGFVKGSALVTTTH 365
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
174 LGLPHNTFWVPWFWFANLSMKAYLGGRIQRTDVLLOSL-----MNEVCTIL 217
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
366 WGYSSRWYIISTLLMSDALKKANLQTVYDLSLWYSREFKSSFDKMSADSSOLDYFNTIL 425
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
218 RTCCGOLYAYDISIPLVYTVVTVAVYSFFLACLI-----GROFLNEN-----KDY 264
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
426 SROHLALLLE-----PDDQKRINLVNTFSHYITGALTQVPGGKDGRLPDGTAWRHEGNY 481
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
265 PGHEMOLVVPVFTILOFLFYM-----GWLKVAEQLINPEGEDDDDFETNWIIDR 313
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
482 PGYSF-----PAFKNASQLIYLLRDTPFPVSGESGWNKKAMVSA-----WIYS- 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
314 NLOVSLLSVDGMH-----QNLPPMERDMY-----NEAAPOPP-----YTAASARSRHS--- 358
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 NPEVG-LPLAGRHPPFNSPLKSAVOGYWYLAWSAKSSSPDKTLASIYLAISDKTQNESTAI 584
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
359 -----FMGSTFNI-----SLKKEDELEMSKEEADTKKESGYSS--- 392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
585 FGETITPASLPQGFYAFNGGAFGHRWQDKWTLKAYTNVWMSSEIYNKDNRYGRYQSHG 644
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
393 -----TIGCEL--GLQPKNY-----HLPKLDL 412
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
645 VAOIVSNGSLSQGVQEQEGWDMNRMQGATTIHLPLKDL 682
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11

```

US-08-428-949A-1
; Sequence 1, Application US/08428949A
; Patent No. 5716617
; GENERAL INFORMATION:
; APPLICANT: Kiran M. Khandke, John Gotto, Ursula Bul
; TITLE OF INVENTION: Compositions of Chondroitinase I and
; TITLE OF INVENTION: Chondroitinase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,949A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Robinson, Joseph R.
; REGISTRATION NUMBER: 33,448
; REFERENCE/DOCKET NUMBER: 0646/1B017-US2

```

us-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

NAME: Robinson, Joseph R.
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 0646/1B01.7-USA
TELEPHONE: 212-527-7783
TELEFAX: 212-753-6327
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 997 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-948A-1

Query Match 3.0%; Score 88.5; DB 1; Length 997;
Best Local Similarity 19.0%; Pred. No. 6.6;
Matches 87; Conservative 62; Mismatches 134; Indels 175; Gaps 23;

QY 55 SSOQQLFEKALYCDYSYLIPIISFVLGFTVTVSVSRWSQYENLPWPDRLMIQVS-SF 113
DB 300 SDBKQLF-----DNYV-----ILGNYTT-----LMFNISRAY 326
QY 114 VEGKDEGRLLRRLIRYAILQVILIRISITSVYKRPFTLHLVLGPMTHGEHKQLOK 173
DB 327 VLEKDPQK-----AQLKQMYLLMT-----KHLDDQGFVKGSAVLTTHH 365
QY 174 LGLPHNTFWPWFVWFWANLSKAYLGGRIKDTVLQSL-----MNEVCTL 217
DB 366 WGYSSRWYISTLLMSDALKEANLQTVYDLSLWYSREFKSPDMKVSADSDLDYFNTL 425
QY 218 RTQCGQLYADWISIPLVYTVQVTVVAVYSPFLACLI-----GRQFLNPN-----KDY 264
DB 426 SRQHLALLLE-----PDQKRLNVTFSHYITGALTQVPPGKGLRDPGTAWRHEGNY 481
QY 265 PGHEMDLVVPVFTILOFLFYM-----GWLKVAEQLINPFGEDDDDDPETNMLIDR 313
DB 482 PGYSF-----PAFKNASQLIYLLRDTFVSVGESGNNLKAMVSA-----WYS- 525
QY 314 NLOVSLLSVDGNH-----QNLPPMERDMYV-----NEAAPQPP-----YTAASARSRHS--- 358
DB 526 NPEVG-LPLAGRHPFNSPLSKVAQYVWLAWSAKSPDKTLASIIYLAISDKTQNESTAI 584
QY 359 -----FMGSTFNI-----SLKEDLELWSKEADTKKESGYSS-- 392
DB 585 FGETITPASLPQGFYAFNGGAFGIHRWQDKMVTLKAYNTNWSSEIYNKONRYGRYOSH 644
QY 393 -----TIGCFLL-----GLOPKNY-----HLPLKDL 412
DB 645 VAQIVNSGSQLSQGYQOEGWDMNRMOGATTIHLPLKDL 682

RESULT 13
US-08-428-946-1
Sequence 1, Application US/08428946
Patent No. 5855883
GENERAL INFORMATION:
APPLICANT: Kiran M. Khandke, John Gatto, Ursula Bul
TITLE OF INVENTION: Method of Disinsection of Vitreous Body from
TITLE OF INVENTION: Neural Retina of the Eye
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby P.C.
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-7513
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7783
TELEFAX: 212-753-6327
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 997 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-948A-1

Query Match 3.0%; Score 88.5; DB 1; Length 997;
Best Local Similarity 19.0%; Pred. No. 6.6;
Matches 87; Conservative 62; Mismatches 134; Indels 175; Gaps 23;

QY 55 SSOQQLFEKALYCDYSYLIPIISFVLGFTVTVSVSRWSQYENLPWPDRLMIQVS-SF 113
DB 300 SDBKQLF-----DNYV-----ILGNYTT-----LMFNISRAY 326
QY 114 VEGKDEGRLLRRLIRYAILQVILIRISITSVYKRPFTLHLVLGPMTHGEHKQLOK 173
DB 327 VLEKDPQK-----AQLKQMYLLMT-----KHLDDQGFVKGSAVLTTHH 365
QY 174 LGLPHNTFWPWFVWFWANLSKAYLGGRIKDTVLQSL-----MNEVCTL 217
DB 366 WGYSSRWYISTLLMSDALKEANLQTVYDLSLWYSREFKSPDMKVSADSDLDYFNTL 425
QY 218 RTQCGQLYADWISIPLVYTVQVTVVAVYSPFLACLI-----GRQFLNPN-----KDY 264
DB 426 SRQHLALLLE-----PDQKRLNVTFSHYITGALTQVPPGKGLRDPGTAWRHEGNY 481
QY 265 PGHEMDLVVPVFTILOFLFYM-----GWLKVAEQLINPFGEDDDDDPETNMLIDR 313
DB 482 PGYSF-----PAFKNASQLIYLLRDTFVSVGESGNNLKAMVSA-----WYS- 525
QY 314 NLOVSLLSVDGNH-----QNLPPMERDMYV-----NEAAPQPP-----YTAASARSRHS--- 358
DB 526 NPEVG-LPLAGRHPFNSPLSKVAQYVWLAWSAKSPDKTLASIIYLAISDKTQNESTAI 584
QY 359 -----FMGSTFNI-----SLKEDLELWSKEADTKKESGYSS-- 392
DB 585 FGETITPASLPQGFYAFNGGAFGIHRWQDKMVTLKAYNTNWSSEIYNKONRYGRYOSH 644
QY 393 -----TIGCFLL-----GLOPKNY-----HLPLKDL 412
DB 645 VAQIVNSGSQLSQGYQOEGWDMNRMOGATTIHLPLKDL 682

RESULT 12
US-08-428-948A-1
Sequence 1, Application US/08428948A
Patent No. 5741692
GENERAL INFORMATION:
APPLICANT: Khandke, Kiran M.
TITLE OF INVENTION: Chondroitinase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby P.C.
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-7513
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428.948A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

US-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04656
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Gordon, Alan M.
 REGISTRATION NUMBER: 30,637
 REFERENCE/DOCKET NUMBER: 32,390-00/PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-831-3244
 TELEFAX: 201-831-3305
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 997 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-04656-1

Query Match 3.0%; Score 88.5; DB 5; Length 997;
 Best Local Similarity 19.0%; Pred. No. 6.6;
 Matches 87; Conservative 62; Mismatches 134; Indels 175; Gaps 23;

QY 55 SSDQQLLFKALYCDYSYIQLIPISFVLGVYTLVSVRWWSQYENLPWDRMIQVUS-SF 113
 DB 300 SQDKQLF-----DNYV-----ILGNYT-----LMFNISRAY 326
 QY 114 VEGKDEGRLLRLRLRYAILGVLLIRLSISTSVYKRFPTLHLVLGFWTHGEHKQLOK 173
 DB 327 VLEKDPQK-----AQLKQWYLLMT-----KHLDDQGFVKGSALVYTHH 365
 QY 174 LGLPHNTFWPWWFANLKMAYLGGRIQRTVLQSL-----MNEVCTL 217
 DB 366 WGYSSRWYIETLMSDALKEANLQTVDSLWYSREFKSPDMKVSADSSLDYFNTL 425
 QY 218 RTQCQLVAYDWISIPVYTVQVTVAVYSPFLACLI-----GRQFLNPN-----KDY 264
 DB 426 SRQHLALLLE-----PDQKRLNLTFSHYITGALTQVPPGGKGLRPGDTAWRHGNY 481
 QY 265 PGHEMDLVVPVFTILQFLFYM-----GWLKVAEQLINPFEGDDDDDFETNWIIDR 313
 DB 482 PGYSF-----PAFKNASQLIYLLRTPFVSGESGWNLLKAMVSA-----WYVS- 525
 QY 314 NLQVSLLSVDGMH-----QNLPPMERDMY-----NEAAPQPP-----YTAASARSRRHS- 358
 DB 526 NPEVG-LPLAGRHPPNSPLKSAQGYWLMASAKSSPDKTSLASIYLAISDKTQNESTAI 584
 QY 359 -----FMGSTFNI-----SLKKEDELSKEEADTDKESGYSS-- 392
 DB 585 FGTTITPASPLOQFYAFNGGAFGIHRWQDKMVTLKAYTNVWSSSEIYNKDNRYGRYQSHG 644
 QY 393 -----TIGCFPL--GLQPKNY-----HLPLKDL 412
 DB 645 VAQIVNSGSQLSQGYQOEGWDMNRMOGATTIHLPLKDL 682

RESULT 15

US-08-233-008A-6
 Sequence 6, Application US/08233008A
 Patent No. 5578480
 GENERAL INFORMATION:
 APPLICANT: Khandke, Kiran M.
 TITLE OF INVENTION: Methods For The Isolation And
 TITLE OF INVENTION: Purification Of The Recombinantly Expressed chondroitinase
 TITLE OF INVENTION: I and II Enzymes From P. vulgaris
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: American Cyanamid Company
 STREET: One Cyanamid Plaza
 CITY: Wayne

APPLICATION NUMBER: US/08/428,946
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Robinson, Joseph R.
 REGISTRATION NUMBER: 33,448
 REFERENCE/DOCKET NUMBER: 0846/1B017-US3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7783
 TELEFAX: 212-753-6237
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 997 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-428-946-1

Query Match 3.0%; Score 88.5; DB 2; Length 997;
 Best Local Similarity 19.0%; Pred. No. 6.6;
 Matches 87; Conservative 62; Mismatches 134; Indels 175; Gaps 23;

QY 55 SSDQQLLFKALYCDYSYIQLIPISFVLGVYTLVSVRWWSQYENLPWDRMIQVUS-SF 113
 DB 300 SQDKQLF-----DNYV-----ILGNYT-----LMFNISRAY 326
 QY 114 VEGKDEGRLLRLRLRYAILGVLLIRLSISTSVYKRFPTLHLVLGFWTHGEHKQLOK 173
 DB 327 VLEKDPQK-----AQLKQWYLLMT-----KHLDDQGFVKGSALVYTHH 365
 QY 174 LGLPHNTFWPWWFANLKMAYLGGRIQRTVLQSL-----MNEVCTL 217
 DB 366 WGYSSRWYIETLMSDALKEANLQTVDSLWYSREFKSPDMKVSADSSLDYFNTL 425
 QY 218 RTQCQLVAYDWISIPVYTVQVTVAVYSPFLACLI-----GRQFLNPN-----KDY 264
 DB 426 SRQHLALLLE-----PDQKRLNLTFSHYITGALTQVPPGGKGLRPGDTAWRHGNY 481
 QY 265 PGHEMDLVVPVFTILQFLFYM-----GWLKVAEQLINPFEGDDDDDFETNWIIDR 313
 DB 482 PGYSF-----PAFKNASQLIYLLRTPFVSGESGWNLLKAMVSA-----WYVS- 525
 QY 314 NLQVSLLSVDGMH-----QNLPPMERDMY-----NEAAPQPP-----YTAASARSRRHS- 358
 DB 526 NPEVG-LPLAGRHPPNSPLKSAQGYWLMASAKSSPDKTSLASIYLAISDKTQNESTAI 584
 QY 359 -----FMGSTFNI-----SLKKEDELSKEEADTDKESGYSS-- 392
 DB 585 FGTTITPASPLOQFYAFNGGAFGIHRWQDKMVTLKAYTNVWSSSEIYNKDNRYGRYQSHG 644
 QY 393 -----TIGCFPL--GLQPKNY-----HLPLKDL 412
 DB 645 VAQIVNSGSQLSQGYQOEGWDMNRMOGATTIHLPLKDL 682

RESULT 14

PCT-US95-04656-1
 Sequence 1, Application PC/TUS9504656
 GENERAL INFORMATION:
 APPLICANT: Khandke, Kiran M.
 TITLE OF INVENTION: A Novel Protein Designated
 TITLE OF INVENTION: Chondroitinase II and its Use with a Protein Designated
 TITLE OF INVENTION: Chondroitinase I to Achieve Complete Vitreal Disinsertion
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: American Cyanamid Company
 STREET: One Cyanamid Plaza
 CITY: Wayne
 STATE: New Jersey
 COUNTRY: U.S.A.
 ZIP: 07470-8426
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

Search completed: July 10, 2003, 12:27:57
Job time : 16.6264 secs

Fri Jul 11 11:45:19 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:24:19 ; Search time 21.0439 Seconds
(without alignments)

3048.258 Million cell updates/sec

Title: US-09-622-964-29

Perfect score: 2944

Sequence: 1 MTITVTKVANRLGFSSSL.....EHAESYPYRDEAGTKPVLVE 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications_AA.*
- 1: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 6: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	641	21.8	314	10	US-09-768-826-54
2	534	18.1	251	10	US-09-768-826-35
3	444	15.1	261	9	US-09-768-783-4
4	232	7.9	99	9	US-09-768-872-429
5	102.5	3.5	1522	9	US-10-225-567A-346
6	96.5	3.3	839	9	US-10-128-823-2
7	96.5	3.3	839	9	US-10-000-823-5
8	96.5	3.3	839	9	US-09-978-303-34
9	96.5	3.3	839	10	US-09-824-258-2
10	96.5	3.3	839	10	US-09-824-258-4
11	96.5	3.3	839	10	US-09-824-258-8
12	96.5	3.3	1863	9	US-09-832-292-29
13	95	3.2	829	10	US-09-764-367A-7
14	94.5	3.2	586	9	US-10-058-513-37
15	94.5	3.2	771	9	US-10-128-714-3280
16	94.5	3.2	905	9	US-10-128-714-8280
17	94.5	3.2	1393	9	US-10-058-513-2
18	93.5	3.2	299	9	US-09-934-455-318
19	92.5	3.1	1232	10	US-09-801-574-46

20	92	3.1	371	10	US-09-750-373-25	Sequence 25, Appl
21	92	3.1	828	9	US-10-198-070-98	Sequence 98, Appl
22	91.5	3.1	407	9	US-10-270-333-186	Sequence 186, App
23	90.5	3.1	664	10	US-09-823-038A-48	Sequence 48, Appl
24	90	3.1	398	9	US-10-162-012-21	Sequence 21, Appl
25	89.5	3.0	398	10	US-09-875-363-2	Sequence 2, Appl
26	89.5	3.0	281	9	US-10-156-761-12667	Sequence 12667, A
27	89.5	3.0	886	9	US-09-981-397A-22	Sequence 22, Appl
28	89.5	3.0	838	9	US-10-137-316-3	Sequence 3, Appl
29	89.5	3.0	838	9	US-09-978-303-2	Sequence 2, Appl
30	89	3.0	699	10	US-09-759-359A-2	Sequence 24, Appl
31	88.5	3.0	679	9	US-09-877-843-24	Sequence 18, Appl
32	88.5	3.0	1163	10	US-09-893-348-18	Sequence 15, Appl
33	87	3.0	411	9	US-10-229-436-15	Sequence 13, Appl
34	87	3.0	693	9	US-10-229-436-13	Sequence 9, Appl
35	87	3.0	1352	10	US-09-736-968A-9	Sequence 9, Appl
36	87	3.0	1352	10	US-09-736-960-9	Sequence 108, App
37	87	3.0	1980	9	US-09-736-968A-108	Sequence 94, Appl
38	87	3.0	1980	10	US-09-736-960-91	Sequence 91, Appl
39	87	3.0	2040	9	US-09-978-244A-27	Sequence 27, Appl
40	87	3.0	2040	9	US-09-925-297-825	Sequence 825, App
41	86.5	2.9	393	10	US-10-072-621-8	Sequence 8, Appl
42	86.5	2.9	2001	9	US-09-925-299-973	Sequence 973, App
43	86.5	2.9	248	9	US-09-925-299-973	Sequence 973, App
44	86	2.9	248	10	US-09-925-299-973	Sequence 973, App
45	86	2.9	248	10	US-09-925-299-973	Sequence 973, App

ALIGNMENTS

RESULT 1

US-09-768-826-54
; Sequence 54, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PFS12P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-54

Query Match	21.8%	Score	641	DB	10	Length	314
Best Local Similarity	53.8%	Pred. No.	1.3e-52				
Matches	129	Conservative	31	Mismatches	62	Indels	18
Gaps	4						
QY	179	NTFWPWFVFWFANLSKMYLGGRIIDTLLQSLMNEVCTLRTOCGOLYAYDWISIPLVYVQ	238				
Db	20	NKYWPCVWFTNLAQAARDGRDIDIALCLLEELNKYRAKCSMLFYHWDISIPLVYVQ	79				
QY	239	VVTVAVSFFLACLGROFLNPKD-----YFGE-----MDLVVVFVTLQLFLF	283				
Db	80	VVTIAVSFFALSIVGRQFVEEAGAAKPKLLKQGPAPALGDPDMYVPLTTLQFFF	139				
QY	284	YMGWLKVAQLINPFGEDDDDPETWIDRNLOVSLSDVGHONLPMERDMYWEAAP	343				
Db	140	YAGWLKVAEQIINPFGEDDDDPETNQLIDRNLOVSLSDVDEMYQNLPPAKQYWDQOP	199				
QY	344	QPPYT-AAARARRHSFMCSTFNISLKEDLWSKEADTDKESGYSST--IGCFGLG	400				
Db	200	QPPYVATAAESLRFSFLGTTNLRMSDDPEQSLQVFAESPGSGRPAPAAQTLLGRFLGV	259				

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0. Version #1.30


```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-823-5

Query Match
3.3%; Score 96.5; DB 9; Length 839;
Best Local Similarity 19.8%; Pred No. 4.3;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

26 GSIYKLLYGEFLVFIPLYSGIRGLYRMVLSDDQQLFEKALYCDSYIQLIPISFVLGFY 85
477 GEILSVLGG-----VYFFPRGIQYFL-----QRRPSMKTLFVDSYSEM--LFFLOSLF 522

86 VTLVVSRRWSQYEN-----LPPDRMLMIQVSSFVEGKDEBGRLLRLRLTYAILG 135
523 MLATVVLVYFSLHKEYVASMVFSALGWTNMLY-----YTRGQQMG-----IYAVMI 569

136 QVILRLSISTSVYKRPFTLHLVLGAFMT-----HGEHKQLQKLGHPHNTFWVPWYMF 188
570 EKMILRDLC-----RFMFVYVFLFGFSTAVVTLIEDGKNDLPSSESTSHR-----615

189 ANLSMKAYLGGIRDTVLLOSMLNEVCTLRTOCGOLYA-----YDWISIPLV 235
616 -----WRGPACRPP---DSSYN---SLYSTCLLEFKFTIGMGDLFTENYDFKAVFI 662

236 YQVVTVAVYSPFLACILGRQPLNPKDYPGHE-----MDLVVPVFTIQLFLFY 284
663 LLLAYVILTYILLNMLIALMGETVNKIAQESKNIKWLQRAITILDTEKSFCLKMKRKA 722

285 MGWLKVAEQLINPFGEDDDDF-----ETNWI-----IDRLNQVSLLS 321
723 SG--KLLQVGYTPDGKDDYRWCFRVDEVNVTWNTNVTNVIINEDPGNCGEVKRTLSFSLRS 780

322 --VDGMQHONLPPMRDMYVNEAAPPPYTAASARRHSFMGST-----FNISLKEDLEL 375
781 SRVSGRH-----WKNFALVLLREASARDRQSAQPEEVYLRQPSGSLKPDAEV 829

376 WSKBEADTDK 385
830 FKSPAASGEK 839

RESULT 8
US-09-978-303-34
; Sequence 34, Application US/09978303
; Publication NO. US20030049728A1
; GENERAL INFORMATION:
; APPLICANT: Caterina, Michael J.
; APPLICANT: Brake, Anthony J.
; TITLE OF INVENTION: Nucleic acid sequences encoding
; TITLE OF INVENTION: capsacin receptor and capsacin receptor-related
; TITLE OF INVENTION: polypeptides and uses thereof
; FILE REFERENCE: UCAL084CON
; CURRENT APPLICATION NUMBER: US/09/978,303
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/235,451
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/072,151
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 08/915,461
; PRIOR FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-303-34

Query Match
3.3%; Score 96.5; DB 9; Length 839;

```

```

Best Local Similarity 19.8%; Pred. No. 4.3;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

26 GSIYKLLYGEFLVFIPLYSGIRGLYRMVLSDDQQLFEKALYCDSYIQLIPISFVLGFY 85
477 GEILSVLGG-----VYFFPRGIQYFL-----QRRPSMKTLFVDSYSEM--LFFLOSLF 522

86 VTLVVSRRWSQYEN-----LPPDRMLMIQVSSFVEGKDEBGRLLRLRLTYAILG 135
523 MLATVVLVYFSLHKEYVASMVFSALGWTNMLY-----YTRGQQMG-----IYAVMI 569

136 QVILRLSISTSVYKRPFTLHLVLGAFMT-----HGEHKQLQKLGHPHNTFWVPWYMF 188
570 EKMILRDLC-----RFMFVYVFLFGFSTAVVTLIEDGKNDLPSSESTSHR-----615

189 ANLSMKAYLGGIRDTVLLOSMLNEVCTLRTOCGOLYA-----YDWISIPLV 235
616 -----WRGPACRPP---DSSYN---SLYSTCLLEFKFTIGMGDLFTENYDFKAVFI 662

236 YQVVTVAVYSPFLACILGRQPLNPKDYPGHE-----MDLVVPVFTIQLFLFY 284
663 LLLAYVILTYILLNMLIALMGETVNKIAQESKNIKWLQRAITILDTEKSFCLKMKRKA 722

285 MGWLKVAEQLINPFGEDDDDF-----ETNWI-----IDRLNQVSLLS 321
723 SG--KLLQVGYTPDGKDDYRWCFRVDEVNVTWNTNVTNVIINEDPGNCGEVKRTLSFSLRS 780

322 --VDGMQHONLPPMRDMYVNEAAPPPYTAASARRHSFMGST-----FNISLKEDLEL 375
781 SRVSGRH-----WKNFALVLLREASARDRQSAQPEEVYLRQPSGSLKPDAEV 829

376 WSKBEADTDK 385
830 FKSPAASGEK 839

RESULT 9
US-09-824-258-2
; Sequence 2, Application US/09824258
; Patent No. US20010047090A1
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID MALCOLM
; APPLICANT: HAYES, PHILIP DAVID
; APPLICANT: MEADOWS, HELEN JANE
; APPLICANT: DAVIS, JOHN BERESFORD
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30075-D1
; CURRENT APPLICATION NUMBER: US/09/824,258
; CURRENT FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: UK 9805137.8
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: UK 9815791.0
; PRIOR FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: UK 9819278.4
; PRIOR FILING DATE: 1998-09-03
; PRIOR APPLICATION NUMBER: US 09/197,636
; PRIOR FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 839
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-824-258-2

Query Match
3.3%; Score 96.5; DB 10; Length 839;
Best Local Similarity 19.8%; Pred No. 4.3;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

26 GSIYKLLYGEFLVFIPLYSGIRGLYRMVLSDDQQLFEKALYCDSYIQLIPISFVLGFY 85
477 GEILSVLGG-----VYFFPRGIQYFL-----QRRPSMKTLFVDSYSEM--LFFLOSLF 522

```

QY 86 VTLVSRWWSQYEN-----LPWDRMLIQVSSFVEGKDEBGRLLRRTLIRYAILG 135
DB 523 MLATVVLVYFSLHKEYVASMVPSLALGWTNMLY-----YTRGFQOMG-----IYAVMI 569
QY 136 QVLLRSISTSVYKRPPTLHLVLGPMT-----HGEHKOLQKLGPHNTFWPWWWF 188
DB 570 EXMILRDLG-----RPMFVYIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
QY 189 ANLSMKAYLGGRIRODTVLLQSLMNEVCTLRTOCGOLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP-----DSSYN-----SLYSTCLELFPKFTIGMGDLFTENTYDFKAVFII 662
QY 236 YQVTVAVYSPFLACLIGRQFLNPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLLAVVILTYILLNMLIALMGETVKNIAQESKNIWKLQRAITILDTEKSFCLKMKRKA 722
QY 285 MGWLKVAEQINPFGEDDDDF-----ETNWI-----IDRLNQSLLS 321
DB 723 SG--KLLQVGYTPDGKDDYRWCFRVDENVNTWNTNGIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMPERDMYNEAAPQPYTAASARRHRHSFMGST-----FNISLKEDLEL 375
DB 781 SRVSGRH-----WKNFALVLLREASARDRQSAQPEVYLRFQSGSLKPEDAEV 829
QY 376 WSKBEADTDK 385
DB 830 FKSPAASGEK 839

RESULT 10

US-09-824-258-4
; Sequence 4, Application US/09824258
; Patent No. US20010047090A1
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID MALCOLM
; APPLICANT: HAYES, PHILIP DAVID
; APPLICANT: MEADOWS, HELEN JANE
; APPLICANT: DAVIS, JOHN BERESFORD
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30075-D1
; CURRENT APPLICATION NUMBER: US/09/824,258
; CURRENT FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: UK 9805137.8
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: UK 9815791.0
; PRIOR FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: UK 9819278.4
; PRIOR FILING DATE: 1998-09-03
; PRIOR APPLICATION NUMBER: US 09/197,636
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 839
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (144)(194)(198)
US-09-824-258-4

Query Match 3.3%; Score 96.5; DB 10; Length 839;

Best Local Similarity 19.8%; Pred. No. 4.3;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

QY 26 GSIYKLLYGEFLVFIPLYISIRGLYRWLVSSDQQLLFEKALYCDYSYIQLIPISFVLGY 85
DB 477 GEILSVLGG-----VYFFPRGIQYFL-----QRRPSMKTLEVDYSYSEM--LFFQLQSLF 522
QY 86 VTLVSRWWSQYEN-----LPWDRMLIQVSSFVEGKDEBGRLLRRTLIRYAILG 135
DB 523 MLATVVLVYFSLHKEYVASMVPSLALGWTNMLY-----YTRGFQOMG-----IYAVMI 569

QY 136 QVLLRSISTSVYKRPPTLHLVLGPMT-----HGEHKOLQKLGPHNTFWPWWWF 188
DB 570 EXMILRDLG-----RPMFVYIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
QY 189 ANLSMKAYLGGRIRODTVLLQSLMNEVCTLRTOCGOLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP-----DSSYN-----SLYSTCLELFPKFTIGMGDLFTENTYDFKAVFII 662
QY 236 YQVTVAVYSPFLACLIGRQFLNPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLLAVVILTYILLNMLIALMGETVKNIAQESKNIWKLQRAITILDTEKSFCLKMKRKA 722
QY 285 MGWLKVAEQINPFGEDDDDF-----ETNWI-----IDRLNQSLLS 321
DB 723 SG--KLLQVGYTPDGKDDYRWCFRVDENVNTWNTNGIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMPERDMYNEAAPQPYTAASARRHRHSFMGST-----FNISLKEDLEL 375
DB 781 SRVSGRH-----WKNFALVLLREASARDRQSAQPEVYLRFQSGSLKPEDAEV 829
QY 376 WSKBEADTDK 385
DB 830 FKSPAASGEK 839

RESULT 11

US-09-824-258-8
; Sequence 8, Application US/09824258
; Patent No. US20010047090A1
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID MALCOLM
; APPLICANT: HAYES, PHILIP DAVID
; APPLICANT: MEADOWS, HELEN JANE
; APPLICANT: DAVIS, JOHN BERESFORD
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30075-D1
; CURRENT APPLICATION NUMBER: US/09/824,258
; CURRENT FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: UK 9805137.8
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: UK 9815791.0
; PRIOR FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: UK 9819278.4
; PRIOR FILING DATE: 1998-09-03
; PRIOR APPLICATION NUMBER: US 09/197,636
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 839
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-824-258-8

Query Match 3.3%; Score 96.5; DB 10; Length 839;

Best Local Similarity 19.8%; Pred. No. 4.3;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

QY 26 GSIYKLLYGEFLVFIPLYISIRGLYRWLVSSDQQLLFEKALYCDYSYIQLIPISFVLGY 85
DB 477 GEILSVLGG-----VYFFPRGIQYFL-----QRRPSMKTLEVDYSYSEM--LFFQLQSLF 522
QY 86 VTLVSRWWSQYEN-----LPWDRMLIQVSSFVEGKDEBGRLLRRTLIRYAILG 135
DB 523 MLATVVLVYFSLHKEYVASMVPSLALGWTNMLY-----YTRGFQOMG-----IYAVMI 569
QY 136 QVLLRSISTSVYKRPPTLHLVLGPMT-----HGEHKOLQKLGPHNTFWPWWWF 188
DB 570 EXMILRDLG-----RPMFVYIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
QY 189 ANLSMKAYLGGRIRODTVLLQSLMNEVCTLRTOCGOLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP-----DSSYN-----SLYSTCLELFPKFTIGMGDLFTENTYDFKAVFII 662

Qy 236 YQVVTAVYGFETACIGROFLNPKDYPGHE-----MDLVVPVFTIQLFLFY 284
Db 663 LLLAVILTYILLNMLIAGETVKNIAQESKNWIKLQRAITITLDTKSLKCKRAFR 722
Qy 285 MGWLKVAEQLNPFGEEDDDF-----ETNWI-----IDRNLOVSLLS 321
Db 723 SG--KLLQVYTPDGKDYRCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLRS 780
Qy 322 --VDGMHQLNPPMERDMYNEAAPPPYTAASARRHSFWGST-----FNISLKEDLEL 375
Db 781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPDAEV 829
Qy 376 WSKKEADTDK 385
Db 830 FKSPAASGEK 839

RESULT 12
US-09-832-292-29
; Sequence 29, Application US/09832292
; Patent No. US2002017705A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey
; TITLE OF INVENTION: MAMMALIAN ALPHA-KINASE PROTEINS, NUCLEIC ACIDS AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: 601-1-098CIP US/09/832,292
; CURRENT APPLICATION NUMBER: 2001-04-10
; PRIOR FILING DATE: 2001-04-10
; PRIOR FILING DATE: 09/632,131
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 29
; LENGTH: 1863
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-832-292-29

Query Match 3.3%; Score 96.5; DB 9; Length 1863;
Best Local Similarity 20.2%; Pred. No. 14;
Matches 129; Conservative 76; Mismatches 179; Indels 255; Gaps 39;

Qy 35 EFLVFTFLY-YSIRGLVRVLSDDQLLFEKALYCDYSIQ-----LIPISFVLGYVTVL 89
Db 886 EWIVIAFTYALEKV-REVFMSEAGKISQIKWFSDFYNSDTIAISFFVGFGLRF- 943
Qy 90 VSRW--WSQYEN-----LPWPDRLMIQVSSFVEGKDEBGRLLRRLTYAI 133
Db 944 GAKWYINAYDNHVFVAGRLIYCLNIIFWYVRL-----DFLAVNQAGPYV----- 990
Qy 134 LGQVLIIRSIYSTVYKRPPTLHLVLVAGPMTHGEHKOLKGLPHNTFWVP-----WVWF 188
Db 991 ---MMIGWVANMFY---IVVIMALVLSF-----GVPRKAILPHEPWSLSA 1033
Qy 189 ANLSMKAY--LGGRIKRTVLQSLMNEVC-----TLRTQCG-----QLYAYDWI 230
Db 1034 KDIVFHPYWMIFGEV-----YAYEIDVCANDSTLPTICPGTWLTPFLQAVLVFYQYII 1087
Qy 231 SIPL-----VYTVQVTVFA-----VYSFPLACLIGRQFLNPKDYPGHEMDLVVPV 276
Db 1088 MVNLLIAFFNNVYLQVRAISIVWKYQRYHEIMAY-----HEKVPVLPPL 1132
Qy 277 TILQ-----FLVGMWLKV---AEOLINP-FGEEDDDDPET- 307
Db 1133 IILSHIVSLFCVCVKRRKDKTSDGPKLFLTEEDQKLUHDEQCVMFYFEKDDKNSG 1192
Qy 308 -----NMIIIDRNLOVSLLSVDGMHQLNPPMERDMYNEA 341
Db 1193 SEERIRVTFERVEQMSIQIKEVGDRVNY-IKESLQ-SLDSQIGHQLDLSALTVDTLTKLT 1250
Qy 342 APQPPYTAASARRHSFWGSTFNISLKK-----EDL-----ELMSKEADTDKESGYS 391

Db 1251 AQK-----ASEASKVNEI--TRELSISKHLAQLNIDDDVPVRLWKKPSAVNTLSSS--- 1300
Qy 392 STTGCEFLGLQPKNYHLPLKDLTKLCSKNPLLEGCKDANKQKQD-VKKFGLDPLK 450
Db 1301 -----LPQGDRE-----SNNPFL---C-NIFMKDEKDPQYNLFGQD-LPV 1335
Qy 451 VPRFK-----RRSHGCPQAPSS-----HPTEQ-----SAPSSSDTGDG 484
Db 1336 IPORKEFNIPEAGSSCCALFPSPVPELQRRHGVEMLKI FNKNQKLGSPNPSPP----- 1391
Qy 485 PSTDYQIECHMKKKTVEFNINIPESPT--EHLQORRLDQ 521
Db 1392 -----HMSPPPKYFSVSTPSQSCSKHLESTTKDQ 1421

RESULT 13
US-09-764-367A-7
; Sequence 7, Application US/09764367A
; Patent No. US20020072101A1
; GENERAL INFORMATION:
; APPLICANT: Gaughan, Glen
; APPLICANT: Ramanathan, Chandra
; TITLE OF INVENTION: NOVEL HUMAN NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING CAT
; TITLE OF INVENTION: CHANNELS
; FILE REFERENCE: 5624,252,999
; CURRENT APPLICATION NUMBER: US/09/764,367A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/177,554
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 7
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-367A-7

Query Match 3.2%; Score 95; DB 10; Length 829;
Best Local Similarity 19.6%; Pred. No. 5.9; Mismatches 53; Indels 132; Gaps 17;
Matches 83; Conservative 53;

Qy 26 GSIYKLLYGEFLVFIIFYYSIR--GLYRMVLSDDQLLFEKALYCDYSIQIPIFVL 82
Db 476 GEILSVLGVYFFFGIOYFLQRRPSMKTLPFDSYSEMLFFLQSLFMLATVVLASNVFSL 535
Qy 83 GPVTLVSRWMSQYENLPWPDRLMIQVSSFVEGKDEBGRLLRRLTYAILGQVLI 142
Db 536 A-----LGWTNMLY-----YTRGFQOMG-----IYAVMIERKILRD 566
Qy 143 ISTSVYKRPPTLHLVLVAGPMTHGEHKOLKGLPHNTFWVPWVWFAVLSMKA 195
Db 567 LC-----RPMFYIIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR----- 605
Qy 196 YLGGRIKRTVLQSLMNEVCITLRTQCGLYA-----YDWISIPLYVTVQVTV 242
Db 606 WRGPACRPP---DSSYN---SLYSTCLELFTKFTIGMGDLFTENYDFKAVFILLIAYVI 659
Qy 243 AVYSFPLACLIGRQFLNPKDYPGHE-----MDLVVPVFTIQLFVGMWLKVA 291
Db 660 LTYILLNMLIALMGETVKNIAQESKNWIKLQRAIFILDTEKSFCLKMRAFRSG--KLL 717
Qy 292 EQLINEPFGEDDDF-----ETNWI-----IDRNLOVSLLS--VDGMH 326
Db 718 QVGYTPDGKDDYRCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLRSRVSGRH 777
Qy 327 QNLPPMERDMYNEAAPPPYTAASARRHSFWGST-----FNISLKEDLEWSKEAD 382
Db 778 -----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPDAEVFKSPAAS 826
Qy 383 TDK 385
Db 827 GEK 829

APPLICANT: Broshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patent in version 3.1
SEQ ID NO 3280
LENGTH: 771
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-10-128-714-3280

Query Match 3.2%; Score 94.5; DB 9; Length 771;
Best Local Similarity 20.6%; Pred. No. 5.9; Indels 149; Gaps 25;
Matches 95; Conservative 61; Mismatches 157

QY 139 ILASISVSVKRPPTLHLVLAGFMTHGKHQKQKLGPHNFWFVWVPMFANLSMKAYLG 198
Db 94 IPRFLDNTSVRV-CLYVMSVWLLVYPEDQPF--LRTAHEIY-----VRYKELTKAIVLA 146
QY 199 GRTRDTVLLQSLMNEVC--TLRTQCGOLYAYD--WISIPLVYTVQVTVAVYSFELACL-- 252
Db 147 IRLNDTOLIKSDINATSDPVLKNQMAFLVARQOIWLDMP-----EEENTSFMOCLNN 199
QY 253 --IGRFLNPKDYPGHEMDLVVP-----VFTILOFLFYMGWLKV-----AEQLINP 297
Db 200 ITISKHFKS-----LGKELDILEPKNPEDYKTHLESSRGAGLTNVD SARHNLASAFVNA 254
QY 298 -----FGEDD-----DDPETNWI-----IDR----- 313
Db 255 FVNAAGFGKDNMLGDDDKGSMWTKDKDGMSTTTASFGMLLSKDTDTGLDRIDKFTYATE 314
QY 314 -NLQVSLLSVDGMHQNLPMPERD-----MYWNEAAPPPVTAASARSRHSFMGST 363
Db 315 EQIKAGALLGTGMITGSGVRLEADPVALLGDPEVWQKRS--VPVRVAIIMGLGLAYAGSN 372
QY 364 FNIISKEDLELWSK--BEADTKKESGYSS-TTG-CFLGLQPKNYHLPLKDKTKLLCS 419
Db 373 -----KELEILLPVVEDVSLDMQLSAMRAVSLGLIFVG---SSNHQVSEAIATTLM-- 422
QY 420 KNPLEGCKDANQKQKQVW-KPKG-----LDFLKCVPRFKRGSHC 461
Db 423 -----DEERQKHLKDKWTRFMALGLALLYFGRQEEVDVILDKAVDH----- 465
QY 462 GPQAPSSHPTQSPSSSDTGDGPSTDYQETCHMKKKTVEFN 503
Db 466 -----PMKPTSVLASVCWAGTGTVVKLQELHLHCNDLIEEN 503

Search completed: July 10, 2003, 12:28:59
Job time : 23.0439 secs

APPLICANT: Eos Biotechnology, Inc.
APPLICANT: Gish, Kurt C.
APPLICANT: Mack, David H.
APPLICANT: Afar, Daniel
TITLE OF INVENTION: Uses of PBH1 in the Diagnosis and Therapeutic Treatment
TITLE OF INVENTION: of Prostate Cancer
FILE REFERENCE: 018501-005910US
CURRENT APPLICATION NUMBER: US/10/058,513
CURRENT FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: US 60/263,951
PRIOR FILING DATE: 2001-01-24
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patent in ver. 2.1
SEQ ID NO 37
LENGTH: 586
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: exons 17-28 of PBH1
US-10-058-513-37

Query Match 3.2%; Score 94.5; DB 9; Length 586;
Best Local Similarity 18.8%; Pred. No. 3.9;
Matches 88; Conservative 55; Mismatches 125; Indels 201; Gaps 23;

QY 42 LYSIRGLRVMVLSDDQQLLFEKALYCDYSYIQLIPISVLGYFTLVVSRMWSQYENLP 101
Db 157 LFVFIAGIVFRLHSSNKSLSYGRVIFCLDY-----IIFTLRIHIFTVSR-----NL- 204
QY 102 WPDRLMIQVSSFVEGKDEGRLLRLIRYAILGQVLLIRLSITSYKRPPTLHLVL-- 159
Db 205 GPKIMLQMTSIEMSSSGSSIPTLRFPEF-----VVLQISIGT-----SSHEVMLS 253
QY 160 -----AGFMTHGKHQKQKLGPHNFWFVWVPMFANLSMKAYL 197
Db 254 DRCVLLPVPLCGVDGLCVARQILRQNEOR-----WRWIFRSVIYEPYL 298
QY 198 G--GRTRDTVLLQSLMNEVC--TLRTQCGOLYAYD--WISIPLVYTVQVTVAVYSFELACL-- 252
Db 299 AMFGQVPSDVGTYDFAHCTFTGNSKPLCVLDEHNLPRPEWITIFLVCIYMLSTNI 358
QY 240 VTVAVYSFFLACLIGRQFLNPKNDYPGHEMDLVVPVFTTILQFLFYMGWLKVAEQLINPFG 299
Db 359 LLVNLVAMFGCVAG-----GLVQVL----- 379
QY 300 EDDDDPETNWIIDNLOVSLLSVDGMHQNLPMPERDWMYNEAAPPPVTAASARSRHSF 359
Db 380 -----ENNLKVS-----QKQKQARELTAKPKYTLAAGFR-- 414
QY 360 MGSTFNISLKKEDLELWSKBEADTKKESGYSTIGCFLGLOP-----KNYHLPLK 410
Db 415 -----WT-----SAVTAC---LQPARCLPPTGTRQGHKISLE 442
QY 411 DLKTKLL-----CS---KNPL-LEGQCKDANQKQKQVW-KPKG-----LDFLKCVPRFKRGSHC 461
Db 443 MHKGKIAEFSGQHQMATGCGQDPFNHLRMGGYTVGTVQSNNDQVWFQ 491

RESULT 15
US-10-128-714-3280
Sequence 3280, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wenqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:04 ; Search time 16.1337 Seconds
(without alignments)
3283.199 Million cell updates/sec

Title: US-09-622-964-29
Perfect score: 2944
Sequence: 1 MTITYTNKVNARLGSFSSL.....EHASYPYRDEAGTKPVLYE 551

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	828.5	28.1	612	2 T32368	hypothetical prote
2	771.5	26.2	557	2 T32367	hypothetical prote
3	756	25.7	499	2 T27630	hypothetical prote
4	755.5	25.7	405	2 T27971	hypothetical prote
5	749.5	25.5	584	2 T19565	hypothetical prote
6	736.5	25.0	632	2 S44917	ZK688.2 protein -
7	722.5	24.5	1355	2 T28715	hypothetical prote
8	665	22.6	413	2 T21644	hypothetical prote
9	663.5	22.5	523	2 T18782	hypothetical prote
10	658.5	22.4	450	2 T18781	hypothetical prote
11	656.5	22.3	405	2 S42371	hypothetical prote
12	655	22.2	513	2 T24210	hypothetical prote
13	646.5	22.0	459	2 S40708	hypothetical prote
14	619.5	21.0	613	2 T16885	hypothetical prote
15	616	20.9	512	2 T19806	hypothetical prote
16	601	20.4	387	2 H89192	protein F32G8.4 [i
17	601	20.4	400	2 T21670	hypothetical prote
18	599.5	20.4	420	2 B88710	hypothetical prote
19	503	17.1	400	2 T20050	protein C43G2.4 [i
20	495	16.8	444	2 T20048	hypothetical prote
21	487.5	16.6	530	2 T28037	hypothetical prote
22	477.5	16.2	806	2 T15468	hypothetical prote
23	466.5	15.8	411	2 T28038	hypothetical prote
24	412	14.0	434	2 T20922	hypothetical prote
25	114.5	3.9	516	2 A80665	probable membrane
26	109.5	3.7	1707	2 T18951	hypothetical prote
27	109.5	3.7	2606	2 T03159	large tegument pro
28	108.5	3.7	315	2 A80677	probable membrane
29	108.5	3.7	560	2 T07964	(S)-N-methylcocciu

30	108	3.7	489	2 S50396	hypothetical prote
31	103.5	3.5	4092	1 S38128	dynein heavy chain
32	102.5	3.5	331	2 AD2063	hypothetical prote
33	102.5	3.5	362	2 T32242	hypothetical prote
34	102.5	3.5	613	2 G82338	conserved hypotet
35	102.5	3.5	1045	2 T18630	hypothetical prote
36	102.5	3.5	1522	2 T00028	brain-specific ang
37	102.5	3.5	2178	2 S29237	calcium channel pr
38	102.5	3.5	2259	2 S29236	calcium channel pr
39	101	3.4	720	2 G64230	stringent response
40	100	3.4	388	2 T16861	hypothetical prote
41	100	3.4	620	2 S55086	probable membrane
42	99.5	3.4	564	2 A48141	ferric reductase (
43	99.5	3.4	570	2 S60426	probable membrane
44	99.5	3.4	685	2 B82297	c-di-GMP phosphodi
45	99.5	3.4	1027	2 B90527	atp-binding protei

ALIGNMENTS

RESULT 1

T32368
hypothetical protein C01B12.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C/Accession: T32368
R/Scheet, P.; Maggi, L.
submitted to the EMBL Data Library, September 1997
A/Description: The sequence of C. elegans cosmid C01B12.
A/Reference number: Z21156
A/Accession: T32368
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-612 <SCH>
A/Cross-references: EMBL:AF025458; PIDN:AAB70976.1; GSPDB:GN00020; CESP:C01B12.3
A/Experimental source: strain Bristol N2; clone C01B12
C/Genetics:
A/Gene: CESP:C01B12.3
A/Map position: 2
A/Introns: 25/3; 60/2; 105/2; 138/3; 212/3; 319/3; 369/2; 467/2; 508/3; 573/1
C/Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 28.1%; Score 828.5; DB 2; Length 612;
Best Local Similarity 38.3%; Pred. NO. 1.7e-59;
Matches 160; Conservative 86; Mismatches 145; Indels 27; Gaps 4;

QY	1	MTITYTNKVNARLGSFSSLLLCWRGSIYKLLYGELFYFIYYSIRGLYRMVLSDDQL	60
DB	1	MTVYSLDVASSPFCLYKLLFRWKSIVKSWAELVVMCLYAVLSVIYRCLLTMKORA	60
QY	61	LFEKLALYCDYIQLIPISFVLGVTVLVSRWWSQVENLPWDRMLQVSVFVGKDBE	120
DB	61	TFEDLCIFFDYFNFIPTFMVGVSVPFRWQIFDNICWIDTPCLWITQYIKETER	120
QY	121	GRLLRLTIRVAILGQVILRSISTSVYKRPPTLHLVLAGFMTHGEHQKLGLPHNT	180
DB	121	AKCVRRNCIRYSILTQAMVYRDVAASVKRPFTFNHLVTAGLTKEKAEFESIPSHAK	180
QY	181	FWVPWFANLSMKAYLGGRIIDRTVLQSLMNEVCTLTQCGQLYAYDWISIPLYTVQV	240
DB	181	YWQPMHMLFSGMITLARDDEGMISDIIYYVDLMKQRQFRVNTLSLTLFQWVPVLYTVQV	240
QY	241	TVAVYSPFLACLIGRQFLNPN----KQVPGHEMDLVVPFTILOFLFYGMKLVKAEQLIN	296
DB	241	HLAVRSFLIALFGRIYHLPESNRLNDFK-OTIDLYVDFIMSLQLFFIFFGMKVAEVLIN	299
QY	297	PFGEDDDDFFETNWIIDRNQLVSLSVDMGHQNLPPMERDMYWNAAPOPPYTAASARGRR	356
DB	300	PLGEDDDDFECNWLDRNLQVGLMVDVYVRYPTLEKQKQWEDAIASEPLVTAESAMRPL	359
QY	357	HSFNGSTFNLSKKED-----LELWSKEADTD-----KKESSYSS 392	

Db 360 NPQVGSADMPTEBEPFVMPVPRRTTLRSMGHWGDMEDTDVFPVVLKHTDRDINSYAS 417

RESULT 2
T32367
hypothetical protein C01B12.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
C:Accession: T32367
R:Sheet, P.; Meggi, L.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid C01B12.
A:Reference number: Z21156
A:Accession: T32367
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-557 <SCH>
A:Cross-references: EMBL:AF025458; PIDN:AAB70978.1; GSPDB:GN000020; CESP:C01B12.5
A:Experimental source: strain Bristol N2; clone C01B12
C:Genetics:
A:Gene: CESP.C01B12.5
A:Map position: 2
A:Introns: 60/2; 105/2; 212/3; 318/3; 368/2; 442/1; 509/3
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 26.2%; Score 771.5; DB 2; Length 557;
Best Local Similarity 31.4%; Pred. No. 6.6e-55;
Matches 171; Conservative 104; Mismatches 132; Indels 77; Gaps 9;

QY 1 MTITYNKVANRLGSPSSLLLCWRGSIYKLLYGEFLVFIFLYYSIRGLYRMVLSQQOL 60
DB 1 MTVAYSIDVATSGLFTQIKVLLRWKGSWKSISWELLWLLCYSLISVIRILLNKAQRE 60
QY 61 LPEKLYALCDSYIQLIPISFVLGFFVYTLVSRWWSQYENLPWDRMLQVSSFVEGDEE 120
DB 61 VFEQLCTFFDTFSFIVPTFMGLFYSIVYRNRWTKVFDNVGWDITSALTIAQYIRGTSEK 120
QY 121 GRLLRRLTIRVAILGQVLLRSISTSVYKRPPTLHLVLGFMTHGHEKQLKGLPHNT 180
DB 121 ARLIRNCVRWIVQVWVFDVSPARRRPTTKHLIGALLTEDELTEFDALTSPOSK 180
QY 181 FWVPVWVFNLSMKAYLGGRIIRDVTLLQSLMNEVCTLRTQCGQLYVDWISIPLYVTQV 240
DB 181 YWQIQIQLFSLVTVAKDEGLIADSVLYVDLIDKMRFPRTKILNLVIFDMVPIPLYVTQV 240
QY 241 TVAVYSFPLACLIGROFLNPKDYPG--HENDLVVPVPTILOFLFYNGWLKVAEQLINP 298
DB 241 NLAVRTYFVLALFGRQFLENNNNIPGAKWKIDIVFPINTSIQIVFVIGWLKVSVMNLPL 300
QY 299 GEDDDDFETNWIIDRNQVLSLLSDGMDHONLPPMERDMYNEAAPQPPYTAASARRHS 358
DB 301 GEDEDEFTNWIIFERNLQGVAVVDQAGRYPTLKDDPFWEDETPQWMDTPTSTRKAHS 360
QY 359 FMGSTFNISLKKEDLELWSKEADTKKESGY----SSTIGCGFLQPKVHLPLKDKLT 414
DB 361 MQGSCINW-----EADLDNGLISYVRRRSRSG-----DDDASSFSTENTRT 403
QY 415 KLLCSKPLLEGQCKDANKQKQVWPKGLDFLKCVPR-----KERGSHCGQAPSS 468
DB 404 QSNASMLP-----RHWQOFRNKISAAIAKFKSEAPPRASVCVVALNIN 449
QY 469 HPTQSA PSSDTGDPSTDYQIEICHMKKTKTFEFLNIPESPTHELQORRLDQSTNQIA 528
DB 450 HRR-----PSTDLREI-----EPLEN-----TDDVVVNVPA 475
QY 529 LMKE 532
DB 476 LVQE 479

RESULT 3
T27630
hypothetical protein ZC518.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000
C:Accession: T27630
R:Thomas, K.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z20396
A:Accession: T27630
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-499 <WIL>
A:Cross-references: EMBL:Z68753; PIDN:CAA92989.1; GSPDB:GN000022; CESP:ZC518.1
A:Experimental source: clone ZC518
C:Genetics:
A:Gene: CESP.ZC518.1
A:Map position: 4
A:Introns: 30/3; 60/2; 106/2; 239/3; 267/3; 315/3; 329/3; 380/2; 405/3; 445/2; 466/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 25.7%; Score 756; DB 2; Length 499;
Best Local Similarity 34.9%; Pred. No. 1.1e-53;
Matches 176; Conservative 87; Mismatches 183; Indels 58; Gaps 10;

QY 1 MTITYNKVANRLGSPSSLLLCWRGSIYKLLYGEFLVFIFLYYSIRGLYRMVLSQQOL 60
DB 1 MTISYTLDSQTNLQSPFSLLRWGSVWKAQVGLAVTAVTAVLLISCIYRYMLSPSQOD 60
QY 61 LPEKLYALCDSYIQLIPISFVLGFFVYTLVSRWWSQYENLPWDRMLQVSSFVEGDEE 119
DB 61 VFEQLIRYFNKLDANIPITFLGFFVSVVVRWGSILNGIGMIDDASLLPATIRGADE 120
QY 120 EGRLLRRLTIRVAILGQVLLRSISTSVYKRPPTLHLVLGFMTHGHEKQLKGLPHN 179
DB 121 ETRVIRENLVYLSQALVLRDISMQVRKFPPTMDTLAASGLMTHBEMDILDHIXDYS 180
QY 180 TFVVPVWVFNLSMKAYLGGRIIRDVTLLQSLMNEVCTLRTQCGQLYVDWISIPLYVTQV 239
DB 181 RYTSIQWNLVTECKKGVDSYLLNKLIVDIBGFRHGLASLLKLDVDPVPLVYQV 240
QY 240 TVAVYSFPLACLIGROFL--NPNKDYPGHMDLVVPVPTILOFLFYNGWLKVAEQLINP 297
DB 241 IFLAVRIYFIMCLIGROFIVTGNPS---GIDLWLPIITTVQVFLVYNGWVKVAEALLNP 296
QY 298 FGEDDDDFETNWIIDRNQVLSLLSDGMDHON-----LPPMERDMYNEAAP 343
DB 297 LGEDDDDELECYIIDKRLITGLSIVDTVMKHDHDTGYSNVEEHMAKTPAQKDFEFGIDKI 356
QY 344 QPPYTAASARRHSFMGSTFNISLKKEDLELWSKEADTKKESGYSSSTIGCGFLQPK 403
DB 357 APLYSMSAERSVHPLVGSASKINLVK-----NKKE-----IVMTPH 393
QY 404 NYHLPLKDLTKLCKSNPLLEGQCKDANKQKQVWPKGLDFL-----KCVPRFRRGS 459
DB 394 KNKLSLDPSEQ---KTYLRRVNVSDHNKAKAQ---RGLERANSPPDKLSKMRSRN 445
QY 460 HCGPQAPSSHPTEQASPSSTGPD 483
DB 446 --GKFRTSANGSQNGVDLWTRAGD 467

RESULT 4
T27971
hypothetical protein ZK675.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T27971
R:Sing, M.
submitted to the EMBL Data Library, November 1994
A:Reference number: Z20448
A:Accession: T27971
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-405 <WIL>
A:Cross-references: EMBL:Z46812; PIDN:CAA86845.1; GSPDB:GN000020; CESP:ZK675.3

A:Experimental source: clone ZK675

C:Genetics:

A:Gene: CESP:ZK675.3

A:Map position: 2

A:Introns: 43/2; 84/2; 110/3; 174/3; 233/3; 264/2; 335/3; 372/1

C:Superfamily: Caenorhabditis elegans hypothetical protein ZK649.5

Query Match 25.7%; Score 755.5; DB 2; Length 405;

Best Local Similarity 41.7%; Pred. No. 8.8e-54;

Matches 150; Conservative 77; Mismatches 118; Indels 15; Gaps 4;

Qy 1 MTIYTNKVNARLGSFSSLLCWRSIYKLYGFLVFIFLYYSIRGLYRMVLSDDOOL 60

Db 1 MTISYSD- - - - - TFLKLLFRWKGSLWKAIRKHLIFLTYMYIINAYYRFGMTKEQN 52

Qy 61 LFEKALYCDYSIOLIPISFVGLPVTVLVSRWMSQYENLWPDRLMIQVSSFVEGKDEE 120

Db 53 EPIKYVMDVGTWKEIPITFLGFIYVAMIVRWWDCCOLIISWPHLLYNVSAIRGQDPE 112

Qy 121 GRLLRRLTIRYAILGQVLILRSISTSVYKRPPTLHLVLVAGFMTHGEHKQLKGL- - - P 177

Db 113 TRIIRKTIARTILTSVLAWRSISLRVLARYPTDHLVDGLTWKEEMVMFKSILVHVDP 172

Qy 178 HNTFWPWWFANLSMKAYLGRIRDTVLLQSLMNEVCTLTQCQOLYAYDWISIPLYVT 237

Db 173 HQKMWPLNIQTMMVRCFEKGTLTHTNELRVLLDALEKRYNGFPQLFIYDWIAIPLVYT 232

Qy 238 QVTVVAVYSFPLACIIGQFLPNKDYPCHEMDLVVPVFTILOFLFYMGWLKVAEOLINP 297

Db 233 QVSTISVYGFALIGQY- - - PSKNEEIEVDVTVFTILOFLFYMGWLKVAEOLINP 290

Qy 298 FGEDDDPFTNWIIDRNQVLSLSDVGMHONLPPMERDMYNEAAPPPYTAASARRRH 357

Db 291 FGADDEDFEYVILERNLEVMILVDELNQVPPVYVESLDEI- - - RLHTSASSKLSNH 348

RESULT 5

T19565

hypothetical protein C29F4.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T19565

R:Kershaw, J.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z19143

A:Accession: T19565

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-584 <WIL>

A:Cross-references: EMBL:Z68335; PIDN:CAA92730.1; GSPDB:GN00022; CESP:C29F4.2

A:Experimental source: clone C29F4

C:Genetics:

A:Gene: CESP:C29F4.2

A:Map position: 4

A:Introns: 36/2; 78/2; 104/2; 149/2; 174/2; 221/2; 256/3; 333/3; 412/2; 484/2; 526/1; 55

C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match.

Best Local Similarity 39.4%; Pred. No. 4.4e-53;

Matches 139; Conservative 81; Mismatches 132; Indels 1; Gaps 1;

Qy 20 LLLCWRSIYKLYGFLVFIFLYYSIRGLYRMVLSDDOQLFEKALYCDYSIOLIPIS 79

Db 64 LIPKWSGLWKAIVLDLIWVCYAFISVIYRYALDRSQDFTFERFMQFCNRRLLDIPIN 123

Qy 80 FVLGFIYTVLVSRWMSQYENLWPDRLMIQVSSFVEGKDEEGLRRLTIRYAILGOVLI 139

Db 124 FMLGFIYTVTVIRNMTQFANLGMIDNIALFTSMISGNDERGRILRSIVRNCVMSQTMV 183

Qy 140 LRSISTSVYKRPPTLHLVLVAGFMTHGEHKQLKGLPHTFWPWWFANLSMKAYLGG 199

Db 184 FDIHIGVKRPPTLETWVAAGIMTSSELKKNEVESRYAKYWLGFNTFNLNEARREG 243

RESULT 7

T28715

hypothetical protein T21D12.9b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T28715

R:Woeschner, J.

submitted to the EMBL Data Library, August 1997

A:Description: The sequence of C. elegans cosmid T21D12.

Qy 200 RIRDTVLLQSLMNEVCTLTQCQOLYAYDWISIPLYTVTVVAVYSGFFLACILIGROFLN 259

Db 244 RIESAYTQNAIAEIRTFERSGLSLITWYDWPPIPLAYPQLVFMALHCHYLVCLVSRQFVI 303

Qy 260 PNKDYPCHEMDLVVPVFTILOFLFYMGWLKVAEOLINPFGEDDDDDPFTNWIIDRNLOVSL 319

Db 304 NSDAVNTTEIDLGVPFMTIIEFIFYMGWLKVAEOLINPFGEDDDDDPFTNWIIDRNLOVSL 363

Qy 320 LSVDMGHONLPPMERDMYNEAAPPPYTAASARRSRHSMFGSTFNISLKKED 372

Db 364 GIVDDSHDDPILKOFMFWNDTV-SPLYSSAAQNRNVNFFGSGATNADAQIPD 415

RESULT 6

S44917

ZK688.2 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

C:Accession: S44917

R:Wilson, R.

submitted to the EMBL Data Library, May 1993

A:Description: Sequence of the C. elegans cosmid ZK688.

A:Reference number: S44913

A:Accession: S44917

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-632 <WIL>

A:Cross-references: EMBL:L16621; NID:g289775; PID:g289780

C:Genetics:

A:Introns: 60/2; 105/2; 268/3; 366/2; 403/2; 435/2; 599/2

Query Match 25.0%; Score 736.5; DB 2; Length 632;

Best Local Similarity 38.6%; Pred. No. 5.7e-52;

Matches 142; Conservative 82; Mismatches 143; Indels 1; Gaps 1;

Qy 1 MTIYTNKVNARLGSFSSLLCWRSIYKLYGFLVFIFLYYSIRGLYRMVLSDDOOL 60

Db 1 MTINYNLAVSTSRKPTWTLFKLLLRWGSIRKAVILELAVLVLYGILSVIYRTALNPGQOR 60

Qy 61 LFEKALYCDYSIOLIPISFVGLPVTVLVSRWMSQYENLWPDRLMIQVSSFVEGKDEE 120

Db 61 TFERIVQCDLSRLSYIPLNFMGLFFVTAVNRWYLYQIIGFIDNIGLMAAEYVGRTEQ 120

Qy 121 GRLLRRLTIRYAILGQVLILRSISTSVYKRPPTLHLVLVAGFMTHGEHKQLKGLPHT 180

Db 121 ARMYRNRIVRYCELAQVLVFRDISMRTTRRPFTLDTVVAAGFMNPKHEDRDEIQYKSK 180

Qy 181 FWPDPVWVFANLSMKAYLGGIRDTVLLQSLMNEVCTLTQCQOLYAYDWISIPLYTVQV 240

Db 181 YWVPFQWAFSLTYEARKKGLIESDYQVVOVDEIKKFTGLAWICNDWDVPIPIPOLV 240

Qy 241 TVAVYSPFLACILIGRQFLNPNKOYPGHEMDLVVPVFTILOFLFYMGWLKVAEOLINPFE 300

Db 241 CLAVHTYFLVCLLARQYVSEHADNKTIEDLYPFIMSTLQFIFFYMGWVKVAEMLNPFGE 300

Qy 301 DDDDFETNWIIDRNQVLSLSDVGMHONLPPMERDMYNEAAPPPYTAASARRSRHSM 360

Db 301 DDDDFECNALIDRNITWVLMWVDOGYDRAPDLKRDDEDFWDEEV-EPLYSEETAKIPNPLK 359

Qy 361 GSTFNISL 368

Db 360 GSVSDVKL 367

QY	1	MTITYNKVANARLGSFSSLLLCWKGSIYKLLIGFVFLVFLFIISLKSIRNA	111
DB	1	MTISYS-----GNVIRILLRWKGSIMRTAWKELLIIYLIVYSRVFVYKGLDIDDD	52
QY	58	Q-----QLLPEKALYCDSYIQLIPISFVLGPFVTVLVSRWWSQYENLPPWLPORLMIQVS	111
DB	53	EDRLKMRMETTCRCQDSYTRLIPIFLGPFVSNVAVRWQFETLWPELILSVLC	112
QY	112	SFVEKGDEGLLRRTLRIRYAILGOVLILRSISVSVYKRPPTLHLVLVAGFMTHGEHKQL	171
DB	113	TVLHQHDEKSKERRHTIARYLNLANALAWRDISSKIRLRFPSVHSLIESGLLTKKYQIL	172
QY	172	QKLGPH-NTFWPWPWFANLSMKAYLGGRIIRDVTLLQSLMNEVCTLRTCCGQLYAYDWI	230
DB	173	EMHAEENSSRWITFLHQLIMRQVBEHPTASLNFQFGEURIFQSLRKULYSYDWMV	232
QY	231	SIPLVTVTVVAVYSFFLACLIGRQFLNPNKDYFGHEMDLVVVPVFTILOFLFYMGWLK	290
DB	233	CVPLVTVVAAALATYSFFFTFLGROPLFPDIE-TGKELDLVVPVFTIVQFLFVGVGPKV	291
QY	291	ABQLNPFGEDEDDPETNWIIDRLNQLVSLSDVGHQON-LPPMER--DMYNE-----	340
DB	292	GQDLMRPFLGDDDDIELNYILDNRVIRISFAIVNOLQESPIPDFESNDDKLWHEMHPPTKD	351
QY	341	-----AAPOPPYTAASAR-----SRHSFM-----GSTEN-----ISLKE	371
DB	352	GETSPIRIPOLPHSKYSKQSEHPPLRHAYVPIDDGKSGIKDLESHHGCVSLSKXD	407
RESULT 9			
T18782			
hypothetical protein B0564.4 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000			
C:Accession: T18782			
R:Lightning, J.			
submitted to the EMBL Data Library, May 1996			
A:Reference number: Z19021			
A:Accession: T18782			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-523 <WIL>			
A:Cross-references: EMBL:Z73422; PIDN:CAA97766.1; GSPDB:GN00022; CESP:B0564.4			
A:Experimental source: clone B0564			
C:Genetics:			
A:Gene: CESP:B0564.4			
A:Map position: 4			
A:Introns: 60/2; 105/2; 320/3; 405/3			
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5			
Query Match 22.5%; Score 663.5; DB 2; Length 523;			
Best Local Similarity 36.4%; Pred. No. 4e-46;			
Matches 139; Conservative 70; Mismatches 157; Indels 13; Gaps 4			
QY	1	MTITYNKVANARLGSFSSLLLCWKGSIYKLLIGFVFLVFLFIISLKSIRNA	60
DB	1	MTINYHKEIKTSHTWKFFVLLPRKSGIMKAIYMETIIFELICVGIISVYVYRTAMSEPSQR	60
QY	61	LFEKALYCDSYIQLIPISFVLGPFVTVLVSRWWSQYENLPPWDLMIQVSSVFGKDE	120
DB	61	TFESVIRYCDRLSFIPLEFVLGFFVTIVVDRWTKLWRTVGFIDDDVCLLANLYVTRTSEK	120
QY	121	GRLLRRTLRIRYAILGOVLILRSISVSVYKRPPTLHLVLVAGFMTHGE-----HKQLQ	172
DB	121	AIYRRIARYCALTQLLVFDVSMTRRRFPPTMETVVAAGFMSKDELDXNSYTTKNS	180
QY	173	KGLPHNTFWPWPWFANLSMKAYLGGRIIRDVTLLQSLMNEVCTLRTCCGQLYAYDWIS	232
DB	181	RLG--KKYIWEANWALCMYTKARKDGYIESDYFKAQMEGEIRTWRTNTEWVCNYDWPVL	237
QY	233	PLVTVTVVAVYSFFLACLIGRQFLNPNKDYFGHEMDLVVVPVFTILOFLFYMGWLKVAE	292
DB	238	PLMTPQLVCLAVNLYFLVLSIIAQ-LVIEKHQWVEVDVYFPVPTFLOFIFYMGWLKVIE	296

Qy	293	QLINPFCEDDDDPFTNWIIDRLQVLSLVDGKHQNLPPMERDMKWNAAPOPPYTAAG	352
Db	297	VMLNPFCEDDDDPFTNAIDRNIITGLKAVDNT-MKTPELLKQDFPEVLVSLLYSESS	355
Qy	353	RSRRHSMFGSGTFNISLKKE	371
Db	356	QISNYHYHGSGTSEVHLEOK	374

RESULT 10
T18781 hypothetical protein B0564_3 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C/Accession: T18781

A;Accession: Z18781
A;Reference number: Z19021
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-450 <WIL>
A;Cross-references: EMBL:Z73422; PIDN:CAA37765.1; GSPDB:GN00022; CESP:B0564.3
A;Experimental source: clone B0564
C;Genetics:
A;Gene: CESP:B0564.3
A;Map position: 4
A;Introns: 60/2; 105/2; 320/3
C;Superfamily: Caenorhabditis elegans hypothetical protein ZK949.5

Query Match 22.4%; Score 658.5; DB 2; Length 450;
Best Local Similarity 34.3%; Pred. No. 8.3e-46;
Matches 141; Conservative 76; Mismatches 165; Indels 29; Gaps 5;

[illegible]

RESULT 11
S42371
hypothetical protein T20G5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S42371
R:Smith, A.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42368

A:Accession: S42371
 A:Name: *Caenorhabditis elegans* hypothetical protein ZK849.5
 A:Molecule type: DNA
 A:Residues: 1-405 <SMI>
 A:Cross-references: EMBL:Z30423; NID:G458479; PID:G458483
 C:Genetics:
 C:Superns: 23/2; 105/2; 267/3; 315/3; 366/2; 404/1
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein ZK849.5

Query Match	22.3%	Score 656.5;	DB 2;	Length 405;
Best Local Similarity	35.2%	Pred. No. 1.1e-45;		
Matches 130; Conservative	85;	Mismatches 151;	Indels 3;	Gaps 3;

Qy	1	MTTYTNKANARIGSFSSLLLCWRGSIYKLLYGEFLVFIIFYYSIRGLYRMVLSSDQOL	60
Db	1	MTVSQNSVATSRTPTWTEFLAIFRWGSGVMSAIWIQYSWMLGYFLVSAIYFILSAYQQQ	60
Qy	61	LEEKALYCDSYQIQIIPISFVLGPYVTVLVSRWSQVENLWPORLMTQVSSFVSGKDBE	120
Db	61	IFVRLVDYNSRMSYVPELDMWLGFFIAGLURFRWLYDIITGFIONIACTATYIRGDSER	120
Qy	121	GRLLRLRIYAILGQVLILRSISTSVYKRPPTLHLHLVLAGFMTHGEHKQLKGLPHNT	180
Db	121	AKQYRRNIIRYCELTQVLIIFRDLMSKARKRPPTLTDVAAGFMPMHEKANFDLIQYNYK	180
Qy	181	FWYPTWFWANLSKAYLGGRIIRDTVLLOSLMNEVCTLRTCQGLYAYYDWISIPLYVTQV	240
Db	181	YFLPFNWAWALVYNARKEGLIEGYYVTVISEDIKKRTGLAWVCNYDWVPLIIPYPTV	240
Qy	241	TVAVYSFELACLIGRQFLNPKOYPGHEMDLVVPVFILOFLFYMGWLVKVAEQLINPGE	300
Db	241	CLAVHMYFFVGILARQYVKSGSEIDP-DMIDLVPFPMTYSIQFVYMGWLVKVGEGLLNPGE	299
Qy	301	DDDDFETNMIIDRNQVSLLSVDGMHQNLPPEMRDYMWNEAAPOPYTAAGARSRR--HSF	359
Db	300	DPDDFETNMIIDRNLANGLKIVDEGYDKTFRLEKDAFWDDTW-VPLIYSEASAHEKRYQR	358
Qy	360	MGSTFNISL	368
Db	359	QGSIAHIKI	367

RESULT 12
T24210
hypothetical protein R13.3 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C/Accession: T24210
R/Matthews, P.
submitted to the EMBL Data Library, May 1996
A/Reference number: Z19855
A/Accession: T24210
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-513 <WILL>
A/Cross-references: EMBL:Z73105; PIDN:CAA97442.1; GSPDB:GN00022; CESP:R13.3
A/Experimental source: clone R13
C/Genetics:
A/Gene: CESP:R13.3
A/Map position: 4
A/Introns: 34/3; 62/2; 107/2; 177/3; 229/3; 261/3; 310/3; 360/2; 429/2; 454/2
C/Superfamily: *Caenorhabditis elegans* hypothetical protein C01B12.5

```

Query Match      22.2%; Score 655; DB 2; Length 513;
Best Local Similarity 31.9%; Pred. No. 1.9e-45;
Matches 166; Conservative 87; Mismatches 205; Indels 62; Gaps 13;

Qy      1  MTITYNKVANKRLGFSFSSLLLCWRGSYKLLVGEFLVFIFYYSRGLYR--MWLSQD 58
      || : : : : : : : : : : : : : : : : : : : : : : : :
Db      1  MTVYNLDVSSASIPSLRLQLRWKGSIMKYLKELFMPIAFITVSSVYRSNLIIGKT 60

Qy      59  QLLPEKALAYCDSYIQLIPISFVLGVFVTVLVSRRWSSQYENLPPWDLMTQVSSFVEKD 118
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      61  RKIWDNFAALFDQMDIFLTLMGLFVTVIRVRWDIFANLGVENTAITVANYIRGTD 120

```

```

Db      339 VAMELLNPFGEADDDFCNCILLIDRNLAIGLTSVDDAYDQLPEVKPDVFVTGSGV--KPLDS 396
QY      350 ASARSRRHFMGGSTFNIS-----LKKEDE-----LELRSK 378
          || : | : ||| : ||| : 
Db      397 DDTRSILKYHF-GSAAQMEEHSIYLKKENKMIAAGKKPNKLKLVWK 440
                                     ||| : 

RESULT 14
Tl6885
hypothetical protein Tl9C3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: Tl6885
R;Fulton, L. The EMBL Data Library, June 1995
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid Tl9C3.
A;Reference number: Z18598
A;Accession: Tl6885
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-613 <FUL>
A;Cross-references: EMBL:U28412; NID:g849235; PID:g849242; PIDN:AAC46599.1; CSP:Tl9C3.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:Tl9C3.1
A;Isotona: 44/3; 95/1; 116/1; 126/2; 160/1; 194/2; 270/3; 347/3; 450/3; 475/1;
```

Query Match	21.0%;	Score	619.5;	DB	2;	Length	613;
Best Local Similarity	36.4%;	Pred. No.	1.9e-42;				
Matches	135;	Conservative	73;	Mismatches	154;	Indels	9;
Gaps	4;						
Qy	1	MTIYTNKANARLGSSPSSLLLCWRGSIYKLLYGBEFLVFIFLYFYISIRGLYRMVLSDDQL	60				
Db	135	MTVSYQDVSNGNPLLFURLGRMKGSIKWSVUGDLFWLLFYAIYFAIRYAFSKQLQT	194				
Qy	61	LFPEKALYCDSPYIQLIPISFVLGPFVTIVVSRWMSQYENLPMPDRLMIQVSSFVEKDEE	120				
Db	195	VFEISIHITDDRMKYLPLTFMGLFPFVTVFERMRGALNVMPITESVALSVAILPGKRE	254				
Qy	121	GRLLRTRIRVAILCOVILIRASISIVSYKRPSTLHLHLVLAGFWTHGEHKOLKGLPHN-	179				
Db	255	DLRTERAIRYVVLHQILVFRDISMRVRRRFPPLKYVNDAGFMRQEBLVLESVNQESSQ	314				
Qy	180	TFWVPEWVFANLSMKAYLGGRIIRDTVLLQSLMNEVCTLRTOCQQLYAYDWISIPLVTVQV	239				
Db	315	TYWVFINWANSIALVAHQKLIQDPTAFNNVIFAKEFRVAMETLIKFDAIPIAIYPOV	374				
Qy	240	VTVAIYSPFLACLIQROGLPNPKDIPGHEMDLVVPVPTLIQILFYGWLKVAQOLINPFG	299				
Db	375	VFLAVRVYFAICLVSRQFLISDMK-SKTQMDWPVPIMTVLEFIFVIGWKMVAEVLNPLG	433				
Qy	300	EDDDDPETNWIIDRNLOVSLLSVDGHQNLPPMRDMWYNEAAPQPPPYTAASARSR--H	357				
Db	434	EDDDDPFNWISIIDNNISIRGMAIVDTTHGYHPLDVDDVF-----SDPNYPAYSSENSQIPR	488				
Qy	358	SFMGSTFNISL	368				
Db	489	NLTGSAAKVEL	499				

```

QY      240 VIVAVISFLAULGKQFLNPNAPDIPGHEHNDV : : : : : : : : : : : : : :
Db      375 VFLAVRVYFAICLVSRQFLISDMK-SKTQMDWVPVIMTVLEFIFVIGNMKVAEVLNPLG 433
QY      300 EDDDDPETWIIDRLNQLVSLSDVGHQNLPPMERDWMYNEAAFPQPPYTAASARSR--H 357
Db      434 EDDDDPEVNSIIDNNSIRGMAIVDTTHGYHPLDVDDVF-----SDPNYLPAYSSENSQIPR 488
QY      358 SFWGSGTFENISL 368
Db      489 NLTGSAAKVEL 499

```

RESULT 15
 T19806
 hypothetical protein C37A5.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T19806
 R:White, S.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19180
 A:Accession: T19806
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-512 <WIL>

Query Match	20.98;	Score 616;	DB 2;	Length 512;
Best Local Similarity	28.38;	Pred. No. 2.9e-42;		
Matches 161;	Conservative 103;	Mismatches 200;	Indels 104;	Gaps 18;
QY	1	MTITYNKVANARILGSPSSLLLCWRGSIYKLLYGFLVFIPLFYISIRGLYRMVSSDQOL	60	
DB	1	MTVSYNLDVSSVSFFNFPKLLFRWRGSVKWSIWSSELVLWLLGYLLVMVYRVALTTEQKA	60	
QY	61	LFEKIALYCDSYI-QLPIGIFVLGFTYTLVVSRWMSQYENLPWDRMLMIQVSSFVEGKDE	119	
DB	61	GVRKYIDHLDQNLKDCVPLTFMLAFFVTIIVDRWKNMFANIGFTIENTAIAIATLVKGTGG	120	
QY	120	EGRLRLTLYRYATILGOVLILRSISTSVYKRFPTLHLVLVLAGFMTHGEHKOLQKGLPHN	179	
DB	121	DVLAKRTIIRYLVTLOVLVFRDLSLKVRRRFPNHDALIKAGFLQ--DHESIILGGDNGR	178	
QY	180	T-FWVPWVWFANLSMKAYLGGRIIRDITVLLQSLMNEVCTLRTOCGQLYAYOWISIPLVYTQ	238	
DB	179	TNYWMPVNWSSAILQKLPEDGNI PAAPL FNSVWQEVKTFRSNMATLCNYDWDVPIAYPQ	238	
QY	239	VVTVAVYSFFLACILIGRQFLNPNDKYDQGHENDLVVPVPTIILOFLFYGMWLKVAQQLNPF	298	
DB	239	VVFAVRVYFPTCLFTRQHLDMEDT---KTIDYVFPILTVPQPTFFGMGMKVAQALLNPL	295	
QY	299	GEDDDDETAWIIDRNQLQVSLSDYDGMHONLPMDRMWYNEAAPQPPYTAASARS-RRH	357	
DB	296	GEDDDDECNLTIDRNATGTHAIVNSKYSVDPEMLADKF-NDPSVAPYYPEKVIDSGADH	354	
QY	358	SFMGSTNISLXKEDLELWSKEEADTDKKEGSGYSTICGFLGLOPKNVHLPKOLKTKLL	417	
DB	355	ALVGSAGQVTL-----ABENDT---IDMMKVDL-379		
QY	418	CSKNPLLEGQCKDANKQKQDWKPKGLDFLKCVPFKERG---SHCGQAPS-SHPTEQ	473	
DB	380	--NSPIVVGR---RNTNNTSTIRR-----RLSSAFGRRRSHSVQHLGPEKPEPTSPFSQ	427	
QY	474	S-AP-----SSSDTGDGPGSDTQGEICHMKKKTVE	501	
DB	428	SMAQRPYGAELSNGLSLGITSQSHLPKLEESSETTD-PLPAQOPVFLPRTLSE	486	
QY	502	FNLNIPESPT-----EHLQORLUDQM	522	
DB	487	---EPTSPTLPFDOTLEHVDNNGLNKM	510	

Search completed: July 10, 2003, 12:27:12
Job time : 18.467 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:20:43 ; Search time 9.11903 Seconds
(without alignments)
2506.127 Million cell updates/sec

Title: US-09-622-964-29

Perfect score: 2944

Sequence: 1 MTITYTNKVNARLGSFSSL.....EHAESYPYRDEAGTKPVLYE 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1908	64.8	585	1 VMD2 HUMAN	O76090 homo sapien
2	756	25.7	499	1 YHDI CAEEL	O23369 caenorhabdi
3	755.5	25.7	405	1 YS63 CAEEL	O09379 caenorhabdi
4	736.5	25.0	632	1 Y022 CAEEL	P34672 caenorhabdi
5	665	22.6	413	1 YV4Q CAEEL	O45435 caenorhabdi
6	663.5	22.5	523	1 Y0E4 CAEEL	Q17529 caenorhabdi
7	658.5	22.4	450	1 Y0E3 CAEEL	Q17528 caenorhabdi
8	656.5	22.3	456	1 YNK4 CAEEL	P34577 caenorhabdi
9	655	22.2	513	1 YXAK CAEEL	Q21973 caenorhabdi
10	646.5	22.0	459	1 YKT8 CAEEL	P34319 caenorhabdi
11	619.5	21.0	501	1 YSV1 CAEEL	Q22566 caenorhabdi
12	601	20.4	400	1 YV6L CAEEL	Q19978 caenorhabdi
13	599.5	20.4	420	1 YCBL CAEEL	Q94175 caenorhabdi
14	546	18.5	884	1 YAVK CAEEL	Q17851 caenorhabdi
15	502.5	17.1	602	1 YHS4 CAEEL	O19303 caenorhabdi
16	466.5	15.8	411	1 YHS5 CAEEL	O18304 caenorhabdi
17	436	14.8	81	1 VMD2 MOUSE	O88870 mus musculu
18	412	14.0	434	1 YSWJ CAEEL	O45363 caenorhabdi
19	108.5	3.7	315	1 YNEE SALTY	Q82706 salmonella
20	108.5	3.7	315	1 YNEE SALTY	Q82706 salmonella
21	108	3.7	489	1 YN20 YEAST	P40210 saccharomyc
22	104.5	3.5	389	1 O95C DROME	Q9VH66 drosophila
23	103.5	3.5	4092	1 DYHC YEAST	P36022 saccharomyc
24	102.5	3.5	613	1 UBID VIBCH	Q9KV48 vibrio chol
25	102.5	3.5	1522	1 BA13 HUMAN	O60242 homo sapien
26	102.5	3.5	2259	1 CCAE RABIT	Q02343 oryctolagus
27	101.5	3.4	306	1 Y114 FALSO	Q8XY11 raistonia s
28	101	3.4	720	1 SPOT MYCCE	P47520 mycoplasma
29	100	3.4	620	1 YN20 YEAST	Q03162 saccharomyc
30	99.5	3.4	564	1 FRP1 SCHPO	Q04800 schizosacch
31	99.5	3.4	570	1 YGR0 YEAST	P53109 saccharomyc
32	98.5	3.3	802	1 YGN9 YEAST	P53121 saccharomyc
33	98	3.3	4447	1 PKSK_BACSU	P40803 bacillus su

34 95.5 3.2 347 1 NU2M HIPAM Q9ZZZ0 hippopotamu
35 94 3.2 487 1 USID_HAEDU Q9RPX1 haemophilus
36 93.5 3.2 299 1 HSF6_ARATH Q98CW4 arabidopsis
37 93 3.2 374 1 Y006_BORBU O51039 borelia bu
38 93 3.2 459 1 NU4M_PELSU O79677 pelomedusa
39 92.5 3.1 199 1 YRF2_SHIFL P37788 shigella fl
40 92.5 3.1 679 1 YKR9_YEAST P34237 saccharomyc
41 92.5 3.1 1663 1 HAPD_HUMAN P34237 saccharomyc
42 92.5 3.1 1862 1 ANK1_MOUSE O60229 homo sapien
43 92 3.1 492 1 Y821_RICPR Q02357 mus musculu
44 91.5 3.1 669 1 YH06_HAENIN Q9ZCD6 rickettsia
45 91 3.1 347 1 NU2M_PIG P45335 haemophilus
O79875 sus scrofa

ALIGNMENTS

RESULT 1
VMD2 HUMAN
ID VMD2 HUMAN STANDARD; PRT; 585 AA.
AC O76090; O75904;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bestrophin (Viteliform macular dystrophy protein). (TUI58).
GN VMD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS BMD.
RX MEDLINE=98367043; PubMed=9700209;
RA Marquardt A., Stoehr H., Passmore L.A., Kraemer F., Rivera A.,
RA Weber B.H.F.;
RT "Mutations in a novel gene, VMD2, encoding a protein of unknown
RT properties cause juvenile-onset vitelliform macular dystrophy (Best's
RT disease).";
RL Hum. Mol. Genet. 7:1517-1525 (1998).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS BMD P-6; H-85; C-93; N-227 AND E-299.
RX MEDLINE=98324772; PubMed=9662395;
RA Petrukhin K., Koisti M.J., Bakall B., Li W., Xie G., Marknell T.,
RA Sandgren O., Forsman K., Holmgren G., Andreasson S., Vujic M.,
RA Bergen A.A.B., McGarty-Dugan V., Figueroa D., Austin C.P.,
RA Metzker M.L., Caskey C.T., Wadelius C.;
RT "Identification of the gene responsible for Best macular dystrophy.";
RL Nat. Genet. 19:241-247 (1998).
RN [3]
RP VARIANTS BMD HIS-13; CYS-93; CYS-218; ASP-300; GLU-301 AND ILE-307.
RX MEDLINE=99265978; PubMed=10331951;
RA Caldwell G.M., Kakuk L.E., Griesinger I.B., Simpson S.A., Nowak N.J.,
RA Small K.W., Maumenee I.H., Rosenfeld P.J., Sieving P.A., Shows T.B.,
RA Ayagari R.;
RT "Bestrophin gene mutations in patients with Best vitelliform macular
RT dystrophy.";
RL Genomics 58:98-101 (1999).
RN [4]
RP VARIANTS BMD V-10; V-82; C-92; H-96; S-135; C-218; S-218 AND K-293.
RX MEDLINE=99320852; PubMed=10394929;
RA Bakall B., Marknell T., Ingvaest S., Koisti M.J., Sandgren O., Li W.,
RA Bergen A.A.B., Andreasson S., Rosenberg T., Petrukhin K., Wadelius C.;
RT "The mutation spectrum of the bestrophin protein -- functional
RT implications.";
RL Hum. Genet. 104:383-389 (1999).
RN [5]
RP VARIANTS AND/BMD, VARIANT AVMD K-146, AND VARIANT BULL'S EYE Q-119.
RX MEDLINE=99381534; PubMed=10453731;
RA Allikmets R., Seddon J.M., Bernstein P.S., Hutchinson A., Atkinson A.,
RA Sharma S., Gerrard B., Li W., Metzker M.L., Wadelius C., Caskey C.T.,
RA Dean M., Petrukhin K.;
RT "Evaluation of the Best disease gene in patients with age-related
RT macular degeneration and other maculopathies.";

Hum. Genet. 104:449-453 (1999).

[6] VARIANTS BMD P-16; C-17; H-92; C-218; H-218; L-235 AND S-296.
MEDLINE=21138457; PubMed=11241846;
Marchant D., Gogart K., Boutboul S., Pequignot M., Sternberg C.,
Dureau P., Roche O., Uteza Y., Hache J.C., Puech B., Puech V.,
Durum V., Mouillon M., Munier P.L., Schorderet D.F., Marsac C.,
Dufier J.L., Abitbol M.;
Identification of novel VMD2 gene mutations in patients with Best
vitelliform macular dystrophy.";
Hum. Mutat. 17:235-235 (2001).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: DEFECTS IN VMD2 ARE THE CAUSE OF BEST MACULAR DYSTROPHY
(BMD); ALSO KNOWN AS VITELLIFORM MACULAR DYSTROPHY TYPE 2. BMD IS
AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY TYPICAL "EGG-YOLK"
MACULAR LESIONS DUE TO ABNORMAL ACCUMULATION OF LIPOFUSCIN WITHIN
AND BENEATH THE RETINAL PIGMENT EPITHELIUM CELLS. PROGRESSION OF
THE DISEASE LEADS TO DESTRUCTION OF THE RETINAL PIGMENT EPITHELIUM
AND VISION LOSS.
CC -!- DISEASE: DEFECTS IN VMD2 COULD BE THE CAUSE OF OTHER FORMS OF
MACULOPATHY AS BULL'S EYE MACULOPATHY, AND ADULT VITELLIFORM
MACULAR DEGENERATION (AMD).
CC -!- DISEASE: IN RARE CASES, DEFECTS IN VMD2 MAY INCREASE
SUSCEPTIBILITY TO AGE-RELATED MACULAR DEGENERATION (AMD).
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -!- DATABASE: NAME=VMD2 mutation database;
WWW="http://www.uni-wuerzburg.de/humanogenetics/vmd2.html".
CC -!- DATABASE: NAME=Mutations of the VMD2 gene;
NOTE=Retina International's Scientific Newsletter;
WWW="http://www.retina-international.com/sci-news/vmd2mut.htm".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).

EMBL; AF073500; AAC64926.1; JOINED.
DR EMBL; AF073491; AAC64926.1; JOINED.
DR EMBL; AF073492; AAC64926.1; JOINED.
DR EMBL; AF073493; AAC64926.1; JOINED.
DR EMBL; AF073494; AAC64926.1; JOINED.
DR EMBL; AF073495; AAC64926.1; JOINED.
DR EMBL; AF073496; AAC64926.1; JOINED.
DR EMBL; AF073497; AAC64926.1; JOINED.
DR EMBL; AF073498; AAC64926.1; JOINED.
DR EMBL; AF073499; AAC64926.1; JOINED.
DR EMBL; AF057169; AAC64343.1; --
DR EMBL; AF057170; AAC64344.1; --
DR EMBL; AF073501; AAC33766.1; --
DR Gene; HGNC:12703; VMD2.
DR MIM: 153700; --
DR InterPro: IPR000615; Worm_fam_8.
DR Pfam: PF01662; DUP289; 1.
DR ProDom: PD002802; Worm_fam_8; 1.
KW Alternative splicing; Disease mutation: Polymorphism; Vision.
FT VARSPLIC 291 483
ASQNLNPFDEDDDETWTIVDNLOVSLAVDEMHDLPRLR
MEFDPMYNKEPFPPTTAASAQAFRRASPSTENISLNKER
MFQPNQDEEADAAGIIRGLFGSDHDPFRANSRYKL
WPKRESLLHGELPKKHAKAQNVRGEEDNKAWLKAVDAFK
SAPLYORPGYSGAPOTPLSPMPFLPELE -> GLSRALLG
WPHGORHQQLLETTRMOCQERKVRSVSQAAMRTVPVPA
TREAEAGESIEPGRRLWMQSSSTPLERMMLRPCTGLST
GICRCPCWLMMWRCTRCLGWSRCTGTGISPSHSPPQLLPPS
SVPEPLWAPPST (IN ISOFORM 2).
T -> P (IN BMD AND AVMD).
/FTId=VAR_000830.
V -> A (IN BMD).
/FTId=VAR_000831.
V -> M (IN BMD)

FT 10 10 VARIANT
FT 10 10 VARIANT
FT 13 13 VARIANT
FT 16 16 VARIANT
FT 17 17 VARIANT
FT 21 21 VARIANT
FT 24 24 VARIANT
FT 25 25 VARIANT
FT 25 25 VARIANT
FT 27 27 VARIANT
FT 58 58 VARIANT
FT 67 67 VARIANT
FT 73 73 VARIANT
FT 82 82 VARIANT
FT 85 85 VARIANT
FT 92 92 VARIANT
FT 92 92 VARIANT
FT 92 92 VARIANT
FT 93 93 VARIANT
FT 96 96 VARIANT
FT 99 99 VARIANT
FT 100 100 VARIANT
FT 104 104 VARIANT
FT 119 119 VARIANT
FT 135 135 VARIANT
FT 141 141 VARIANT
FT 146 146 VARIANT
FT 209 209 VARIANT
FT 216 216 VARIANT
FT 218 218 VARIANT
FT 218 218 VARIANT
FT 218 218 VARIANT
FT 218 218 VARIANT
FT 224 224 VARIANT
FT 227 227 VARIANT
FT 227 227 VARIANT
FT 227 227 VARIANT

/FTId=VAR_000832.
A -> T (IN BMD).
/FTId=VAR_000833.
A -> V (IN BMD).
/FTId=VAR_010468.
R -> H (IN BMD).
/FTId=VAR_010469.
S -> F (IN BMD).
/FTId=VAR_010470.
F -> C (IN BMD).
/FTId=VAR_010471.
L -> V (IN BMD).
/FTId=VAR_000834.
W -> C (IN BMD).
/FTId=VAR_000835.
R -> Q (IN BMD).
/FTId=VAR_000836.
R -> W (IN BMD).
/FTId=VAR_000837.
S -> R (IN BMD).
/FTId=VAR_000838.
Q -> L (IN BMD).
/FTId=VAR_000839.
L -> V.
/FTId=VAR_000840.
I -> N (IN BMD).
/FTId=VAR_010472.
L -> V (IN BMD).
/FTId=VAR_010473.
Y -> H (IN BMD).
/FTId=VAR_000841.
R -> C (IN BMD).
/FTId=VAR_010474.
R -> H (IN BMD).
/FTId=VAR_010475.
R -> S (IN BMD).
/FTId=VAR_000842.
W -> C (IN BMD).
/FTId=VAR_000843.
Q -> H (IN BMD).
/FTId=VAR_010476.
N -> K (IN BMD).
/FTId=VAR_000844.
L -> R (IN BMD).
/FTId=VAR_000845.
D -> E (IN BMD).
/FTId=VAR_000846.
E -> Q (IN BULL'S EYE MACULOPATHY).
/FTId=VAR_010477.
G -> S (IN BMD).
/FTId=VAR_010478.
R -> H (IN BMD).
/FTId=VAR_000847.
A -> K (IN AVMD).
/FTId=VAR_010479.
S -> N (IN BMD).
/FTId=VAR_000848.
T -> I (IN AMD; SPORADIC).
/FTId=VAR_010480.
R -> C (IN BMD).
/FTId=VAR_000849.
R -> H (IN BMD).
/FTId=VAR_010481.
R -> Q (IN BMD).
/FTId=VAR_000850.
R -> S (IN BMD).
/FTId=VAR_000851.
L -> M (IN BMD).
/FTId=VAR_000852.
Y -> C (IN BMD).
/FTId=VAR_000853.
Y -> N (IN BMD).
/FTId=VAR_000854.

Query Match 64.8%; Score 1908; DB 1; Length 585;
 Best Local Similarity 64.5%; Pred. No. 4e-147;
 Matches 379; Conservative 59; Mismatches 100; Indels 50; Gaps 10;

QY 1 MTITITKVNARLGSFSLLCWRSIYKLYLGRFLVFIFLYYSIRGLYRMVLSDDOOL 60
 DB 1 MTITITSQVANARLGSFSLLCWRSIYKLYLGRFLVFIFLYYSIRGLYRMVLSDDOOL 60

QY 61 LPEKALYCDYIQLIPISFVLGFYVTLVSVRWMSQYENLPWPDRLMIQVSSFVEGKDE 120
 DB 61 MEKLTLYCDYIQLIPISFVLGFYVTLVSVRWMSQYENLPWPDRLMIQVSSFVEGKDE 120

QY 121 GRLLRRTLIYAILGOVILRLSISVYKRPFTLHLVLVLAGFMTHGEHKQLKGLPHNT 180
 DB 121 GRLLRRTLIYAILGOVILRLSISVYKRPFTLHLVLVLAGFMTHGEHKQLKGLPHNT 180

QY 181 FWPVWVFWANLGMKAVLGRIRDTVLQSLMNEVCTLRTOCGQLYAYDWISIPLVYTOV 240
 DB 181 FWPVWVFWANLGMKAVLGRIRDTVLQSLMNEVCTLRTOCGQLYAYDWISIPLVYTOV 240

QY 241 TVAVYSFFLACILGRQFLNPKNDYPGHEMDLVVPVFTILQFLFYMGWLKVAEQLINPGE 300
 DB 241 TVAVYSFFLACILGRQFLNPKNDYPGHEMDLVVPVFTILQFLFYMGWLKVAEQLINPGE 300

QY 301 DDDDPETNWIIDRNLOVLSLSDVGMHONLPMDRMWNEAAPQPPYTAASARSRRHSPM 360
 DB 301 DDDDPETNWIIDRNLOVLSLSDVGMHONLPMDRMWNEAAPQPPYTAASARSRRHSPM 360

QY 361 GSTFNISLKEDELWSKEADTDKESGYSSTIGCFGLQPKNHYLPLKDLTKLLCSK 420
 DB 361 GSTFNISLKEDELWSKEADTDKESGYSSTIGCFGLQPKNHYLPLKDLTKLLCSK 420

QY 421 NPLL--EGCKD-----ANQNKQD--VWKFGLDPLKCVPRFKRGSGCGPAPSS--- 468
 DB 421 NPLL--EGCKD-----ANQNKQD--VWKFGLDPLKCVPRFKRGSGCGPAPSS--- 468

QY 469 --HPTQSPSS--SDTG-----DPSDYOICIMKKKT 499
 DB 469 --HPTQSPSS--SDTG-----DPSDYOICIMKKKT 499

QY 500 VEFNL--NIPESPTLQORLDQMSNIOALMKEHAEYS---PYRDEA 543
 DB 500 VEFNL--NIPESPTLQORLDQMSNIOALMKEHAEYS---PYRDEA 543

QY 537 VEFNLTDMEIPEINHLKE-PLQSPNTIHTLKDHPDTPWALENDEA 583
 DB 537 VEFNLTDMEIPEINHLKE-PLQSPNTIHTLKDHPDTPWALENDEA 583

RESULT 2
 YHDI CAEEL STANDARD; PRT; 499 AA.

AC Q23369;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZC518.1 in chromosome IV.
 GN ZC518.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Thomas K.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z68753; CAA92989.1; -.
 DR WormPep; ZC518.1; CE06601.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 499 AA; 57229 MW; 4E87725437A5C9DF CRC64;

Query Match 25.7%; Score 756; DB 1; Length 499;
 Best Local Similarity 34.9%; Pred. No. 9.7e-54;
 Matches 176; Conservative 87; Mismatches 183; Indels 58; Gaps 10;

QY 1 MTITITKVNARLGSFSLLCWRSIYKLYLGRFLVFIFLYYSIRGLYRMVLSDDOOL 60
 DB 1 MTITITSQVANARLGSFSLLCWRSIYKLYLGRFLVFIFLYYSIRGLYRMVLSDDOOL 60

QY 61 LPEKALYCDYIQLIPISFVLGFYVTLVSVRWMSQYENLPWPDRLMIQVSSFVEGKDE 119
 DB 61 VPEQLIRYFNKLDANIPLTLLGFFSVFVARWGSILNGIWDASLLFATYIRGADE 120

QY 120 EGRLLRRTLIYAILGOVILRLSISVYKRPFTLHLVLVLAGFMTHGEHKQLKGLPHN 179
 DB 121 ETRVIRRNLYRVLSQALVLRDISMQVRKRPFTMDTLAASGLMTHSEMDILDHKDPYS 180

QY 180 TFWVWVFWANLGMKAVLGRIRDTVLQSLMNEVCTLRTOCGQLYAYDWISIPLVYTOV 239
 DB 181 RYMTSIQMSLNLVTECKGKGVSYLLMNKIVDEIGKFRHGLASLLKYDWPVPLVYPQV 240

QY 240 TVAVYSFFLACILGRQFL--NPNKDYPGHEMDLVVPVFTILQFLFYMGWLKVAEQLINP 297
 DB 241 IFLAVRIYFMICLGRQFIVTGNPS-----GIDLWLPITTVQVFLVYMGWKVAEQLINP 296

QY 298 FGEDDDPETNWIIDRNLOVLSLSDVGMHON-----LPPMERDMWNEAAP 343
 DB 297 LGEDDDLECNVYIIDKNLITGLSIVDTWVKHDDTGYSMVEBHMKTPAQKDEPWGIDKI 356

QY 344 QPPYTAASARSRRHSPMSSTFNISLKEDELWSKEADTDKESGYSSTIGCFGLQPK 403
 DB 357 APLYSMESAERSVHPLVGSASKINLVK-----NKXE-----IVMTTPH 393

QY 404 NVHLPLKDLTKLCSKNPLLEGCKDANQKNQKDVWKFGLDPL-----KCVPRFKRGS 459
 DB 394 KNKSELDPSEQ-----KTYLRRVNVSDHNAKHAQO-----RGLERANSYDKLCKVSRSRN 445

QY 460 HCGPQAPSSHPTEQSAFSSSDTG 483
 DB 446 --GKFRTSANGSQNGVDLWTRAGD 467

RESULT 3
 YS63 CAEEL STANDARD; PRT; 405 AA.

AC Q09379;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK675.3 in chromosome II.
 GN ZK675.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Sims M.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 281074; CAB03043.1; -
DR WormPep; F32B6.9; CR09864.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1_fam_8.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 413 AA; 48965 MW; A66E69A83C78790B CRC64;

Query Match 22.6%; Score 665; DB 1; Length 413;
Best Local Similarity 35.1%; Pred. No. 1.8e-46;
Matches 146; Conservative 86; Mismatches 130; Indels 54; Gaps 11;
1 MTITYNKNVANARLGSFSSLLCMRGSYKLLYGBFLVFIPLYSYIRGLYRM---VLSSD 57
1 MTISYS-----GNVIRILLRWKGSWRTAKELLILYISRVFYLGIDUDD 52
58 Q-----QLLFEKALYCDYQILIPISFVLGFFVYTLVVSWMQYENLPPDRMLQVS 111
53 EDDLKMRMPETFCRCQDSVTLPLTLFLGFFYVNVVAVRWRQFELYWPEDILSVLC 112
112 SFVEGKDEGRLRLTIRYAILGOVLIRLSISTSVYKRPPTLHLVLGFWTHGEKQL 171
113 TVLHDEKSKRRHTTARYLANALAWRDISSKIRLRFPSVHSLIESGLLTKKEYQIL 172
172 QKGLPLH-NTFWVWVWFANLSKAYLGGRIKRDVTLQSLMNEVCTLRTOCGQLYADWI 230
173 EMHAENESSRWITPLHWIQLIMQVEEHEKPTASLNFQVGLRIFRQSLKGLSYDWMV 232
231 SIPLVYTVVAVYSFFLACLGROFLNPNKDPGHEMDLVVPVFTTLQFLFYMGWLKV 290
233 CVPLVYTVVAVATYSFFFTLGRQPLFDIE-TGKELDLVVPVFTIVQFLFFVGFVKV 291
291 AEQLINPFGEDDDDFETNWIIDRLNQLSLLSDGMHQN-LPMPER--DMYVNE----- 340
292 GDLMRPFGLDDDDIELNYILDRNVRISFAVNLQESPIPDFESNDKLMWHMHPPTKD 351
341 -----AAPPPVTAASAR-----SRHSEF-----GSTFN-----ISLKE 371
352 GETSPIRIFQLPHSKYSKQLSEHPHPRHAYVIPPDDGKSGIKOLESHHGCVSLKDD 407

RESULT 6
YQ84 CAEEL
ID YQ84 CAEEL STANDARD; PRT; 523 AA.
AC Q17529;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein B0564.4 in chromosome IV.
GN B0564.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lightning J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 273422; CAA97766.1; -
DR WormPep; B0564.4; CR05178.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1_fam_8.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 523 AA; 61725 MW; EC3707253E5F9A6 CRC64;

Query Match 22.5%; Score 663.5; DB 1; Length 523;
Best Local Similarity 36.7%; Pred. No. 3.3e-46;
Matches 139; Conservative 70; Mismatches 157; Indels 13; Gaps 4;
1 MTITYNKNVANARLGSFSSLLCMRGSYKLLYGBFLVFIPLYSYIRGLYRMVLSSDQQL 60
1 MTINYHKEIKTSHTWKPFVLLFRWKSISWKAIMETIIFLICYGIIISVVYRTAMSPSQ 60
61 LFEKALYCDYQILIPISFVLGFFVYTLVVSWMQYENLPPDRMLQVSSFVEGKDEE 120
61 TFSVIRYCDRLSPLEFVLGFFVYTLVVSWMQYENLPPDRMLQVSSFVEGKDEE 120
121 GRLRLTIRYAILGOVLIRLSISTSVYKRPPTLHLVLGFWTHGE-----HKQLQ 172
121 AIIYRNRIARYCALTLQVLRDVSMTTRRRPPTMETVVAAGFMSKDELDLYNSYTTKNS 180
173 KGLPLHNTFWVWVWFANLSKAYLGGRIKRDVTLQSLMNEVCTLRTOCGQLYADWISI 232
181 RLG---KYWTPANWALCMYTKARKDGYESDYFKAQMEGEIRTWRTIEWCNDVWVPL 237
233 PLVYTVVAVYSFFLACLGROFLNPNKDPGHEMDLVVPVFTTLQFLFYMGWLKVAE 292
238 PLVYTVVAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVY 296
293 QLINPFGEDDDDFETNWIIDRLNQLSLLSDGMHQN-LPMPERDMYVNEAAPPPVTAASA 352
297 VMLNPFGEEDDDDFETNWIIDRLNQLSLLSDGMHQN-LPMPERDMYVNEAAPPPVTAASA 355
353 RSRRHSEFSGTFSNLSLKE 371
356 QISNYHVGSTSEVHLEQK 374

RESULT 7
YQ83 CAEEL
ID YQ83 CAEEL STANDARD; PRT; 450 AA.
AC Q17528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein B0564.3 in chromosome IV.
GN B0564.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lightning J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; 273422; CAA97766.1; -

```
DR WormPep; B0564.3.; CE05177.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 450 AA; 53275 MW; A0FED9AA76166AD7 CRC64;

Query Match      22.4%; Score 658.5; DB 1; Length 450;
Best Local Similarity 34.3%; Pred. No. 6.9e-46;
Matches 141; Conservative 76; Mismatches 165; Indels 29; Gaps 5;

QY 1 MTITYNKVANARLGSFSSLLLCWRGSIYKLLYGEFLVFLFYISIRGLYRMVLSDDQOL 60
DB 1 MTINYHKHETSHPTWTFLLFKWGSIKWAVYMETIIFLCYIGIISVIKTAGESSOR 60
QY 61 LFEKALYCDYSYIOLIPISFVLGYVTLVSRWMSQYENLPWDRMLIQVSSFEVGEDEE 120
DB 61 TFESLVRYFDKRLSYIPLFVLGPFVTVVNRWTKLYQTIGFDNVGLMANCYIRGATEK 120
QY 121 GRLRLRLIRYAILGQVLIIRLSISTSVYKRPPTLHLVLVLAGFMTHGE-----HKOLQ 172
DB 121 ARIYRRNIMRYCELVQILVFRDMSMTRRRPPTMETVVAAGFMKHELELYNSYDTKYNS 180
QY 173 KGLPHNTFWPVPWFANLSMKAYLGGRIKRTDVLLOSIMNEVCTLTCCGOLYAYDWISI 232
DB 181 KLG---TKYWIIPANWALCMTYKARKDGYIESDYFKAQMEGEIRTWRTNIEWVCNDWVPL 237
QY 233 PLVYTQVTVVAVYFFLAACLGROFLNPNKDYPGHEMDLVVVFVFTILOFLFYMGWLKVAE 292
DB 238 PLMPYQLVCLAVNLVFLVSIARQ-LVTEKHVMDEVDVYFPVMTFLOFIFYNGWLKVID 296
QY 293 QLINPFGEDDDFETNWIIDRLNQLVSLSDVGMHQNLPMPERDMYVNEAAPQPPYTAASA 352
DB 297 VMLNPFGEEDDDFETNALIDRNITMGLMIADN-PNSTELKDKDFYDEVDPVPLYSEBS 355
QY 353 RSRHSPMGSTFNISLKED-----LELWSKEEADYTKKE 387
DB 356 NIPNHHYGSVSEVLEQKGNAPVMMPHSQSAANLRMRMGFSKSVDEDEKD 406

RESULT 8
YXAK_CABEL STANDARD; PRT; 456 AA.
ID YXAK_CABEL
AC P34577.
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein T20G5.4 in chromosome III.
GN T20G5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berka M., Smith A., Lloyd C.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN (2)
RP REVISIONS.
RA Durbin R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z30423; CAA83005.2; -.
DR WormPep; B0564.3.; CE05177.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 456 AA; 53136 MW; EDE22B541307F3C8 CRC64;

Query Match      22.3%; Score 656.5; DB 1; Length 456;
Best Local Similarity 35.2%; Pred. No. 1e-45;
Matches 130; Conservative 85; Mismatches 151; Indels 3; Gaps 3;

QY 1 MTITYNKVANARLGSFSSLLLCWRGSIYKLLYGEFLVFLFYISIRGLYRMVLSDDQOL 60
DB 1 MTVSYNQSVATSRPMTFTLALIFRWGRSVMWSAIQYVWMLGLFLVSAIRFLISAYQQQ 60
QY 61 LFEKALYCDYSYIOLIPISFVLGYVTLVSRWMSQYENLPWDRMLIQVSSFEVGEDEE 120
DB 61 IFVRLVDYVNSRMSYVPLDWMGLGFFIAGVLRFPWLYDIIGFDNIACSTATVIRGDSER 120
QY 121 GRLRLRLIRYAILGQVLIIRLSISTSVYKRPPTLHLVLVLAGFMTHGEHKLQKLGPHNT 180
DB 121 AKQYRRNIRYCELTQVLIIFRDLMSKARKRPPTLDTVAAGFMMPHEKAFNFDLIQYNYNK 180
QY 181 FWPVWTFANLSMKAYLGGRIKRTDVLLOSIMNEVCTLTCCGOLYAYDWISIPLVTVQV 240
DB 181 YFLPFWANALVYNARKEGLIEGDIYVTVISEDIKKFRTGLAVWCYNDWVPLIIPYPTIV 240
QY 241 TVAVYFFLAACLGROFLNPNKDYPGHEMDLVVVFVFTILOFLFYMGWLKVAEOLINPFE 300
DB 241 CLAVHMYFFVGIARQYVKGSEIDP-DMIDLVPFMTSICQFVFMGWLKVGEGLLNPWE 299
QY 301 DDDDFETNWIIDRLNQLVSLSDVGMHQNLPMPERDMYVNEAAPQPPYTAASARR-HSF 359
DB 300 DPDDFETNMLIDRLNGLXIVDEGYDKTPLEKDAFWDDTW-VPLYSEASAEKRYHQ 358
QY 360 MGSTFNISL 368
DB 359 QGSLAHIKI 367

RESULT 9
YXAK_CABEL STANDARD; PRT; 513 AA.
ID YXAK_CABEL
AC Q21973.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein R13.3 in chromosome IV.
GN R13.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Matthews P.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z73105; CAA97442.1; -.
DR WormPep; R13.3; CR06320.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
```


2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans." ;
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; Z29094; CAA82342.1; --
DR PIR; S40708; S40708.
DR WormPep; C07A9.8; CE00497.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 459 AA; 53570 MW; C2D8F69E4078BF9F CRC64;
Query Match 22.2%; Score 646.5; DB 1; Length 459;
Best Local Similarity 33.1%; Pred. No. 6.7e-45;
Matches 134; Conservative 87; Mismatches 149; Indels 35; Gaps 7;
QY 1 MTITNTKVNANRGLSFSSLLLCWRGSIYKLYGFLVFIFLYYSIRGLYRMVLSSDQ- 59
DB 44 LSNVNYDLATSKSLMIVRMIFKRWGSLMQAVYKELIWIWICAYSLVSYRFLTRSQKE 103
QY 60 -----LLPEKALYCDSYIQLIPISFVLGYYVTVVSRWMSQYENLPWDRMLIQVSSFV 114
DB 104 QNKETIIFRFEYCDARMGYLPLNFVLGFCNIIIRRWLKYTLGNDINIALFVSAV 163
QY 115 EKGDEGLRLRLTRLYAILGQVILRLSRISTSVYKFPFLHLVLGTFHGHQLOKL 174
DB 164 RTGDDRARQIRNRIIRYCVISQCLVFRDIHVGRFRFPTLEAVAQAGIMLPEHELEKFSI 223
QY 175 GLPHNTFWPWWFANL-----SMKAYLGRRTRDTVLQSLMNEVCTLTQCGQVAYDM 229
DB 224 KSRQKYVWVFNWALELLNVAKTEKSIDGNARN-----AIAQEISKFSALTIVSMYDM 278
QY 230 ISIPLYTVQVTVAVYSPFLACLIQGRFLNPNKQYPGHEMDLVVPVFTILOFLFYGMWLK 289
DB 279 VPILMYPOLVNMVHTYFFLCIFTRQFFISADAHNKTEDVDLYIPMTIIEFIFYGMWLK 338
QY 290 VAEQLINPFGEDEDDFETNWIIDRLQVLSLLSDVGMHQNLPMERDMYNEAAPPPYTA 349
DB 339 VAMELLNPFGEDEDDFDCNLLIDRLNLAIGLTSVDDAYDQLPQKVPKDPVGTGGSV--KPLDS 396
QY 350 ASARRRHSFMGSTFNIS-----LKKED-----LELWSK 378
DB 397 DTRSLKXHF-GSAAQMEIEISYLKKEENKMIAGKPNKLKUVK 440
RESULT 11
ID YSVL CAEEL STANDARD; PRT; 501 AA.
AC Q22566;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein T19C3.1 in chromosome III.
GN T19C3.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

```

[2]
RN REVISIONS.
RP Durbin R.;
RR Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RL -!- SIMILARITY: BELONGS TO THE BESTORPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 272509; CAA96648.2; -
CC WormPep; F32G8.4; CE23700.
CC InterPro; IPR000615; Worm_fam_8.
CC Pfam; PF01062; DUF289; 1.
CC ProDom; PD002802; Worm_fam_8; 1.
CC Hypothetical protein.
KW SEQUENCE 400 AA; 46750 MW; B1E8ABEC3B062E84 CRC64;
SQ
Query Match 20.4%; Score 601; DB 1; Length 400;
Best Local Similarity 34.9%; Pred. No. 2.7e-41;
Matches 134; Conservative 66; Mismatches 150; Indels 34; Gaps 9;
QY 1 MTITYNKVANRLGFSFSLLCWRGSIYKLLYGEFLVFIFYISIRGLYRMVLSDDQL 60
DB 1 MTISYDEE-----FSSLMRLWRGSIWKAVLKDLIGFYIYIVLAFQWYLLDEKGE 52
QY 61 LPEKALYCDSYQLIPISVLGPFYTVLVSRWWSQVENLPWDRLMIOVSSFVEGKDEE 120
DB 53 YFTGWTMWEIGAQYIPLSFLGFLFFVSLIVARWWEQNCISWPDQXWIMVSACLPG--NE 110
QY 121 GLRLRRLTRYAILGQVLIILRSISTSYKRPFTLHLHLVLGFTMGHGHKQLKGLPHNT 180
DB 111 NMVVRQTIAWSSLSQAIAWSGVSVTKLRFPTERHNVASKLMTTEESYDLYMTNDAPHGK 170
QY 181 FWTPWFWFANLSMKVYLGGRIRTVLLQSLMNEVCTLRTCCGQLYADWTISPLVYTVQV 240
DB 171 WFPIPLWVNLIKKQKQKG-IIDSQMDMLLKQVYTRDGFAMLFVYDWIKIPLVYTVQV 229
QY 241 TVAVYSFPLACLIGRQPLNPKDYPGHEMDLV--PVFTILQFLFYNGWLKVASQLINPP 298
DB 230 AIATYGVFFICLGRQ---PKLDORSMEKETILFPITFTFQMLFYLGWLKVGQ-----F 281
QY 299 GEODDDFETWIIDRNQVLSLVSDGHQHLNLPMERDMYNEAAPQPPYTAAS----- 351
DB 282 SIRKISSELNVLDNRNTAIAHMAASELSDLQPSIGAPM-----VPAVPHTRASFQIDVI 336
QY 352 ARSRHSPMGSTFNISL-KKEDLE 374
DB 337 PKSHLAGFKLSEAEMLKIKPEDLE 360
DB
RESULT 13
YCEL_CABEL STANDARD; PRT; 420 AA.
AC Q94175;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C43G2.4 in chromosome IV.
GN C43G2.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fullon B., Wohldmann P.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

```

```

[2]
RRP REVISIONS.
RRP Watson R.;
RRR submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28412; AAL56625.1; -.
DR WormPep; T19C3.1; CE30182.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
KW SEQUENCE 501 AA; 57930 MW; 1C56C4C70707FC62 CRC64;
SQ
Query Match 21.0%; Score 619.5; DB 1; Length 501;
Best Local Similarity 36.4%; Pred. No. 1.2e-42;
Matches 135; Conservative 73; Mismatches 154; Indels 9; Gaps 4;
QY 1 MTITYTNKVANARLGSPSSILLCHWRGSIYKLLYGEPLVFIFLYYSIRGLYRMVLSSDQQL 60
DB 23 MTVSQYQLDVSSGNPLLFRLRLGLWRGSIKWSVVGDLFWLLFYAIYFAYRFAFSKQLQT 82
QY 61 LFEKIALYCDSYQILQIPISFVLGPFYVTLVWSRMWSQYENLPDRLMIQVSSFVEGKDE 120
DB 83 VFEISHTDHRMKYLLPLTFMGLFVTVFVERWSALNVMFFIESVALSVAVLLPGKRE 142
QY 121 GLLRLRTLRIYALGQVLLIIRSIYSTSVYKRPPTLHLVLGAFMTHGEKQKLGPHN- 179
DB 143 DRLTRRAIRIVVLHQILVLPFRDISMRVRRREPFTLKYYVVDAGFMREQEELDVLESVNQESSQ 202
QY 180 TFWVPWTFANLSMKVNLGGRIQDVTLLQSLMNEVCTLRTOCGQLYAYDWISIPLVTTQV 239
DB 203 TYWVPINWNSALVAHQOKLIDQTPANNVIFAKEFRVAMEYLKEDAPIPIAYPQV 262
QY 240 VTVAVYGFACLIQROFNPNKDPYGHMDLVVPVFTTLOFLFYMGMLKVABOLINPFG 299
DB 263 VFLAVRVYFAICLVSRQFLISDMK-SKTQMDWPVPIMTVLSFIFVIGWMKVAEVLNPLG 321
QY 300 EDDDDFETNWIIDRLNQLVSLSDVGMHQNLPMEERDMYNEAAPOPPYTAASASRR--H 357
DB 322 EDDDDFVNSIIDNNISRGMAIVDTTHGYHPLVDVDF-----SDPNYLPAYSENSQIPR 376
QY 358 SPMGSTFTNISL 368
DB 377 NLTGSAAKVEL 387
RESULT 12
YV6L_CAEBL STANDARD; PRT; 400 AA.
AC QV6L_CAEBL
AC Q19578;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein F32G8.4 in chromosome V.
GN F32G8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RR McMurray A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

```

--1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

This SWISS-PROM entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstations at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; US0069; AAB37559.2; -
CC WormPep; C09B9.3; CE29571.
CC InterPro; IPR00615; Worm_fam_8.
CC Pfam; PF01662; DUF289; 2;
CC ProDom; PD002802; Worm_fam_8; 3.
CC KW Hypothetical protein; Transmembrane; Repeat.
CC TRANSMEM 265 285 POTENTIAL.
CC FT TRANSMEM 398 418 POTENTIAL.
CC SQ SEQUENCE 884 AA; 102851 MW; 8DB83F9699E8FALD CRC64;
CC -----
Query Match 18.5%; Score 546; DB 1; Length 884;
Best Local Similarity 32.8%; Pred. NO. 2.3e-36;
Matches 125; Conservative 80; Mismatches 156; Indels 22; Gaps 10;
QY 1 MTITYNKVANRLGFSILLRCWRSIYKLYAGEFLVFIFYYSIRGLYRMVLSSDQQL 60
DB 1 MTISYYDVATESYFGFKVLFWRKGSVMKLIHRELFMVLVLYYTVLAIYR-TLDEERKK 59
QY 61 LFEKALYCDSYIQLIP--ISFVLGYVTLVSRWWSQVENLPWDRMLIQVSSFVEGKD 118
DB 60 IFRS--NIEHFNIPFSSILTMFLSEFVTTIVQRWNNVFTNMGFIENAAAYAVSFMK-NG 115
QY 119 EGRGLRRTLRVAILGQVLILRSISTSVYKRPPTLHLVLACFMTHGHEKQLKGLPH 178
DB 116 EDVRRQRVIRVLVASQLWKSISIKALRRPNTESITAGFLTKBESTIIONTDSY 175
QY 179 NTPFWPWWFANLSMKAYLGGRIRDTVLQLQSLMNEYCTLRTCQGOLYAYDWISIPLYYTQ 238
DB 176 DSSCVPIRWAIQVLRHQYRSGNPFSSHVSYVRATWKEVSDPETHLSRVKVDWVPIPLAYPQ 235
QY 239 VVTVAVYSPLACLIGROFLNPKNQYPGHEMDLVVPVFTILOFLFYGMGLKVAEQLINPF 298
DB 236 VIFPAVLRYFVICAFKAQYFDLDDDDARVVIHYFFIVTFVQFICLGMGLKVAEALLNPL 295
QY 299 GEDDDDFETNWIIDRLNQVSLLSVDCGMHQLNPMERDMYNEAAPPPYTAASARSR--- 355
DB 296 GEDDDDFEYNFLDSNIYMT-----MSYNC-DVATDSYFNPFKVN-PFTKSYKNNIIF 347
QY 356 --RHSPMGSTFNISKKEDELEW 376
DB 348 QILFRWKGYSWK-SIWKE-LALW 368
CC -----
RESULT 15
YHS4_CABEL
ID YHS4_CABEL STANDARD; PRT; 602 AA.
AC O18303;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK849.4 in chromosome 1.
OS ZK849.4.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RC Kershaw J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RL [2]

```

RP REVISIONS.
RA Durbin R.; (OCT-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted 10/01/2000; (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z82095; CAB05027.2; --
DR WormFep; ZK849.4; CE25696.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 2.
KW Hypothetical protein.
SQ SEQUENCE 602 AA; 67607 MW; 458AB78802BD63A CRC64;

Query Match 17.1%; Score 502.5; DB 1; Length 602;
Best Local Similarity 29.9%; Pred. No. 4.7e-33;
Matches 135; Conservative 72; Mismatches 144; Indels 101; Gaps 12;

Qy 1 MTITYTKVANARLGSSLLLCWRGSIYKLLYGELFVIFLYYYSIRGLYRMVLSDDQL 60
Db 14 MTVTYNRAVSTESIHNFVSIMSYWHGSLVKSINKEIWIIVLYHLILFIRVMP----- 68

Qy 61 LFEKALYCDSYIQ-----LIPISFVLGFYVTVLVSRRWSQYENLPMPLRLMIQVSS 112
Db 69 IP-GWADYCKKVIENWTPHQDMTIPLEFLGFFVTVIDRWKAFQNIPIEICAFVAA 127

Qy 113 FVEG-----KDEGR----- 122
Db 128 AIPGRIRIINKONEVRAPGGQDLRDVPAVEVNHVPVQRPAPAAALPMGAPERPAVAL 187

Qy 123 -----LLRRTLRVAILGQVILRLRSISTVYKRFPTLHHLVLG 161
Db 188 PMGAPEGAIIQPSVDVKLTARTIIRYLVLSQILFREISTYKRFVDLKLVDK 247

Qy 162 FMTHGEHKLOK-LGLPH-NTFWPFWVWFANLSMKAYLGGRIRDTVLQSLMNEVTLRT 219
Db 248 FLTDEELKILSKVKCDHYDVSFLPINWAFSILQE-----HKVKSNEPFINAWNVIRDWQV 303

Qy 220 QCGQLYAVDWISIPLVYTVTVTVAVYFELACLIGROPINPNKDYPGHEMDLVVPVFTIL 279
Db 304 KLSLLRNGDFIPIPLAYPQAVFLAIRPYFLVCLFTRHLDMDDK---KSIDYFFPLMTSL 360

Qy 280 QFLYMGWGLKVAEQILNPPGDDDDPETNWIIDRLNQLVSLSDV---GMHQNLPPMERDM 336
Db 361 QPFIIVGWMKVAEILNPPGDDDDFELNNIIDKLYGLAIVDTECGKH---PEIVKDT 417

Qy 337 YWNEAAPQPPYPTASARSRRHSFMGSTFNISL 368
Db 418 ICKDCLPFYPQN-DDDNEERNALVGSTKINL 448

Search completed: July 10, 2003, 12:24:47
Job time : 11.119 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:21:29 ; Search time 30.163 Seconds
(without alignments)
3763.954 Million cell updates/sec

Title: US-09-622-964-29
Perfect score: 2944
Sequence: 1 MTITYTNKANRLGFSLSL.....EHAESYPYDEAGTPELYE 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB-seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1155	39.2	428	Q8WNR7	Q8wmr7 sus scrofa
2	1078.5	36.6	466	Q9NXP0	Q9nxp0 homo sapien
3	1069.5	36.3	465	Q8VCM0	Q8vcm0 mus musculu
4	961	32.6	721	Q9V3J6	Q9v3j6 drosophila
5	858.5	29.2	535	Q9VDM7	Q9vdm7 drosophila
6	828.5	28.1	612	Q17206	Q17206 caenorhabdi
7	771.5	26.2	557	Q17205	Q17205 caenorhabdi
8	752	25.5	327	Q9VDM6	Q9vdm6 drosophila
9	749.5	25.5	584	Q18303	Q18303 caenorhabdi
10	722.5	24.5	1447	Q16779	Q16779 caenorhabdi
11	714.5	24.3	731	Q9VDM4	Q9vdm4 drosophila
12	637	21.6	551	Q9NA59	Q9na59 caenorhabdi
13	616	20.9	512	Q62095	Q62095 caenorhabdi
14	559	19.0	551	Q965X4	Q965x4 caenorhabdi
15	503	17.1	400	Q17674	Q17674 caenorhabdi
16	495	16.8	444	Q17671	Q17671 caenorhabdi

17	168	5.7	72	4	Q9BR80	Q9br80 homo sapien
18	167.5	5.7	98	5	O01649	O01649 drosophila
19	117.5	4.0	516	16	Q8ZP99	Q8zpp9 salmonella
20	114.5	3.9	516	16	Q8Z772	Q8z772 salmonella
21	110.5	3.8	1139	4	Q9P278	Q9p278 homo sapien
22	109.5	3.7	1707	5	Q17652	Q17652 caenorhabdi
23	109.5	3.7	2606	12	O36414	O36414 alcelaphine
24	108.5	3.7	560	10	O64901	O64901 eschscholzi
25	108	3.7	1311	6	Q9BDH6	Q9bdh6 equus caball
26	102.5	3.5	331	16	Q8YVCO	Q8yvco anabaena sp
27	102.5	3.5	362	5	O17030	O17030 caenorhabdi
28	102.5	3.5	553	5	Q17414	Q17414 caenorhabdi
29	101	3.4	255	5	Q9NAE5	Q9nae5 caenorhabdi
30	101	3.4	503	8	Q8WKM3	Q8wkm3 hylotelephi
31	100.5	3.4	532	10	Q9LIN5	Q9lin5 arabidopsis
32	100	3.4	388	5	Q22454	Q22454 caenorhabdi
33	100	3.4	439	16	Q8XNR9	Q8xnr9 clostridium
34	99.5	3.4	685	16	Q9KU65	Q9ku65 vibrio chol
35	99.5	3.4	854	10	Q9SXH8	Q9sxh8 brassica ol
36	99.5	3.4	1027	16	Q9SR88	Q9sr88 mycoplasma
37	98.5	3.3	518	3	O74283	O74283 coprinus ci
38	98	3.3	467	16	Q9XAR1	Q9xar1 streptomyce
39	97.5	3.3	533	17	Q96Z47	Q96z47 sulfobus
40	97.5	3.3	1349	5	Q9NSD3	Q9nsd3 caenorhabdi
41	97.5	3.3	1413	5	Q9NBD3	Q9nbd3 caenorhabdi
42	97	3.3	616	8	Q9G859	Q9g859 malawimonas
43	97	3.3	642	5	Q9V4V6	Q9v4v6 drosophila
44	97	3.3	856	5	Q963T4	Q963t4 drosophila
45	96.5	3.3	445	10	Q08700	Q08700 brassica na

ALIGNMENTS

RESULT 1

Q8WNR7	PRELIMINARY;	PRT;	428 AA.
ID	Q8WNR7		
AC	Q8WNR7;		
DT	01-MAR-2002 (Tremblrel. 20, Created)		
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Bestrophin (fragment).		
OS	Sus scrofa (pig).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
OX	NCBI_TaxID=9823;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Marmorstein L.Y., McLaughlin P.J., Stanton B., Yan L., Crabb J.W.,		
RA	Marmorstein A.D.;		
RT	"Bestrophin interacts with the b-catalytic subunit of protein		
RT	phosphatase 2A";		
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY064707; AAL40882.1; -		
DR	InterPro; IPR000615; Worm_fam_8.		
DR	Pfam; PF01062; DUF289; 1_		
DR	ProDom; PD002802; Worm_fam_8; 1.		
FT	NON TER 1		
FT	NON TER 1		
SQ	SEQUENCE 428 AA; 48498 MW; 3E6591B2BCD6B383 CRC64;		
Query Match			
Best Local Similarity 55.0%; Score 1155; DB 6; Length 428;			
Matches 241; Conservative 48; Mismatches 85; Indels 64; Gaps 12;			
QY	156	HLVLGFWTHGHHKQKQLGLPHTFPWPVWVFANLSMKAYLGGRIKRDIVLLQSLMNEVC	215
Db	3	HLVKGAFMTPESEHKKLEKLSLPHNSFWPWPVWVFANLSMKAYLGGRIKRDIVLLQSLMNEVC	62
QY	216	TLRTCGQLYADWISLIPLVYTVVTVVAVYSFFLACLGQFLNPKVYFGHEMDLVVPV	275
Db	63	TLRTCGHLIYADWISLVPLVTVVTVVAVYSFFLACLGQFLNPKVYFGHEMDLVVPV	122
QY	276	FTILOFLYMGWLKVAEQILNPFGEDDDDFTNWIIDRLNQVLSLLSDVCGHQLPMPERD	335

```

Db 123 FTFLQFFYAGLWVAEQLINPFGEEDDDFETNWIIVDRSLQVLSAVDEMHDLPPMERD 182
QY 336 MYWNEAAPQPPYTAASARSRRHSFMGSTFNISLKEDLELWKEADTDKKESGYSSTIG 395
Db 183 MYWNPDPHPYTAASQAQRSPSPFGSTFNISLKGDEMEPOPEEE-----EAAHTGILG 236
QY 396 CFLGLQPKNYHLPLKDLTKLLCSKNPLLEGQCKDANOKN-----QKD--VWKF 443
Db 237 HFLGLQSSDHQPPRTNSKTLWPKK---EGHFEGHPKNLARGALDSSQDESKANREG 293
QY 444 GLDLKCVPRKRGSGCGPOAPSSH-----PTQSQP-----SSSDTGDGPST 487
Db 294 G----FSAALCGRPGYHSAPOTPLGHTPMVPPPEESAPLGLRRVSGIDEAAKQSLQPAT 350
QY 488 -----DYQIECHMKKKTVEPNL-NIPESPTEHLQORRLDQSTWIOA 528
Db 351 PSYKSFELIPESAEAEPIQGGSHVRKTVFNLADLSEAP-EHLKEPNLEP-PMGIHA 408
QY 529 LMKHEAESY---PYRDEA 543
Db 409 ILKDRDPYWALENRDEA 426

```

RESULT 2

```

Q9NXP0 ID Q9NXP0 PRELIMINARY; PRT; 466 AA.
AC Q9NXP0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE CDNA FLJ20132 fis, clone COL06441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000139; BAA90970.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 466 AA; 52008 MW; BB9608363A09C2EC CRC64;

```

```

Query Match 36.6%; Score 1078.5; DB 4; Length 466;
Best Local Similarity 60.1%; Pred. No. 9.9e-85;
Matches 199; Conservative 57; Mismatches 74; Indels 1; Gaps 1;

```

```

QY 45 SINGLRMVLSSDQQLFEKLALYCDSYIQLIPISVLGFGYVTLVSRWWSQYENLPWD 104
Db 2 ALSAAYRFLVTEGQKRYFEKLVICDYQYASLIPVSVLGFYVTLVWHRWNQYLCMLPLD 61
QY 105 RLMIQVSSFVEGKDEEGRLLRRLIRYAILGOVLILRSISTSVYKRPPTLHLVLGAFMT 164
Db 62 ALMCVAGTVHGDRDLRRLRLMYAGLSAVLILRSVSTAVFKRPTTDHVVVEAGFMT 121
QY 165 HGEHKQLQKLGPHNTFWPWVFWFANLSMAYLGGRIIRDVTLLQSLMNEVCTILRTCGQL 224
Db 122 REERKKFENLSSYNKYVPCWVFNLSAAQARREGIRDSNLSALKLLLELVNFRKCGML 181
QY 225 YADWISIPLVYTVVTVVAVYSFELACLIIGRQFLPNKDYPGHEMDLVVPVFTILOFLFY 284
Db 182 FHYDWSIVPLVYTVVTVVAVYSFELACLIIGRQFLPNKDYPGHEMDLVVPVFTILOFLFY 241
QY 285 MGWLKVAEQILNPFGEEDDDFETNWIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 344
Db 242 AGWLKVAEQILNPFGEEDDDFETNFIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 301

```

```

QY 345 PPYTAASA-RSRRHSFMGSTFNISLKEDLE 374
Db 302 APYTAATVFLQRPSPFGSTFDITLAKEDMQ 332

```

RESULT 3

```

Q8VCM0 ID Q8VCM0 PRELIMINARY; PRT; 465 AA.
AC Q8VCM0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ20132.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019528; AAH19528.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 465 AA; 51998 MW; E7145E2F893B03F6 CRC64;

```

```

Query Match 36.3%; Score 1069.5; DB 11; Length 465;
Best Local Similarity 58.7%; Pred. No. 5.9e-84;
Matches 199; Conservative 59; Mismatches 80; Indels 1; Gaps 1;

```

```

QY 45 SINGLRMVLSSDQQLFEKLALYCDSYIQLIPISVLGFGYVTLVSRWWSQYENLPWD 104
Db 2 ALSAAYRFLVTEGQKRYFEKLVICDYQYASLIPVSVLGFYVTLVWHRWNQYLCMLPLD 61
QY 105 RLMIQVSSFVEGKDEEGRLLRRLIRYAILGOVLILRSISTSVYKRPPTLHLVLGAFMT 164
Db 62 ALMCVAGTVHGDRDLRRLRLMYAGLSAVLILRSVSTAVFKRPTTDHVVVEAGFMT 121
QY 165 HGEHKQLQKLGPHNTFWPWVFWFANLSMAYLGGRIIRDVTLLQSLMNEVCTILRTCGQL 224
Db 122 REERKKFENLSSYNKYVPCWVFNLSAAQARREGIRDSNLSALKLLLELVNFRKCGML 181
QY 225 YADWISIPLVYTVVTVVAVYSFELACLIIGRQFLPNKDYPGHEMDLVVPVFTILOFLFY 284
Db 182 FHYDWSIVPLVYTVVTVVAVYSFELACLIIGRQFLPNKDYPGHEMDLVVPVFTILOFLFY 241
QY 285 MGWLKVAEQILNPFGEEDDDFETNWIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 344
Db 242 AGWLKVAEQILNPFGEEDDDFETNFIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 301
QY 345 PPYTAASA-RSRRHSFMGSTFNISLKEDLELWKEAD 382
Db 302 APYTAATAFLQRPSPFGSTFDITLAKEDMQFORLDGVD 340

```

RESULT 4

```

Q9V3J6 ID Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG6264 protein (LP05915P).
GN BEST OR CG6264.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

```

Query Match . 32.6%; Score 961; DB 5; Length 721;
Best Local Similarity 50.4%; Pred. No. 2.5e-74;
Matches 19; Conservative 70; Mismatches 108; Indels 10; Gaps 6

Qy 1 MTTTNTKVNAR-LGSPSSLLLCWRGSIYKLLYGEFLVFIYYSIRGLRMVLSSDQ 59
db 1 MTTTITGEVATCRGFGGFLKLLLEWRGSIYKLVMLDLDAFTIYVAINNVYFGLNPAQK 60

```


RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*";
 RT Science 287:2185-2195(2000).
 DR EMBL; AB003531; AAF49648.1; --
 DR FlyBase; FBgn0036492; CG12327.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 SQ SEQUENCE 535 AA; 62710 MW; 7C933A0E5068DF2F CRC64;

Query Match 29.2%; Score 858.5; DB 5; Length 535;
 Best Local Similarity 44.0%; Pred. No. 1.2e-65;
 Matches 165; Conservative 74; Mismatches 129; Indels 7; Gaps 4;

QY 1 MTITYTNKVN-ARLGFSLLLCWRGSIYKLYGFLVPIFLYISIRGLYRMVLSDDQ 59
 DB 1 MTVSYTAEVATCSHFCKWLLMRWRASIYKIIWVLLAFLSCFYFMAYIYVALRDVDK 60

QY 60 LPEKALYCDYIQLIPISFVLGVVTVLVSRWMSQYENLPWDRMTQVSSFVGKDE 119
 DB 61 PVFEDIVMYCHSYNLIPISFVLGVVTVLVSRWMSQYENLPWDRMTQVSSFVGKDE 120

QY 120 EGRLLRRLIRYAILGQVILRLSISVYKRPPTLHLVLGFWTHGKHQKQLKGL--- 176
 DB 121 HGRLLRRLIRYAILGQVILRLSISVYKRPPTLHLVLGFWTHGKHQKQLKGL--- 180

QY 177 PNTFWPVPWVPANLSMKAYLGRIRDTVLQSLMNEVCTLRQCQGLYAYDWISIPLVY 236
 DB 181 KHPKYWPIVWASIVTRAKRGIWDDPSLKSMDLANKFRAGCNMLIHYDISIPLVY 240

QY 237 TQVTVAVYSFFLACIGROFLNPNKDYPGHENDLVVVPVFTILOFLYMGWLKVAEQLIN 296
 DB 241 TQVTVAVYSFFLACIGROFLNPNKDYPGHENDLVVVPVFTILOFLYMGWLKVAEQLIN 300

QY 297 PFGEDDDDFETNWIIDRNLOVSLLSVDGMHONLPMERDMYNEAAP-QPPYTAASARR 355
 DB 301 PFGEDDDDFELNWLIDRNLOVSLLSVDGMHONLPMERDMYNEAAP-QPPYTAASARR 360

QY 356 RHSPMGSTFNISLKK 370
 DB 361 HPE--ASTARLGIPK 373

RESULT 6
 ID O17206 PRELIMINARY; PRT; 612 AA.
 AC O17206;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE C01B12.3 protein.
 GN C01B12.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierzy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.
 elegans*";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Scheet P., Maggi L.;
 RT "The sequence of *C. elegans* cosmid C01B12.1";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF025458; AAB70976.1; --
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 SQ SEQUENCE 612 AA; 71031 MW; DFBB43916541DD44 CRC64;

Query Match 28.1%; Score 828.5; DB 5; Length 612;
 Best Local Similarity 38.3%; Pred. No. 5.6e-63;
 Matches 160; Conservative 86; Mismatches 145; Indels 27; Gaps 4;

QY 1 MTITYTNKVN-ARLGFSLLLCWRGSIYKLYGFLVPIFLYISIRGLYRMVLSDDQ 60
 DB 1 MTVSYTAEVATCSHFCKWLLMRWRASIYKIIWVLLAFLSCFYFMAYIYVALRDVDK 60

QY 61 LPEKALYCDYIQLIPISFVLGVVTVLVSRWMSQYENLPWDRMTQVSSFVGKDE 120
 DB 61 TFEELCIPFTYSNPIPTFMLGFFYVSATFTRWQIFDNIGWIDTTPCLMTQVKGTER 120

QY 121 GRLLRRLIRYAILGQVILRLSISVYKRPPTLHLVLGFWTHGKHQKQLKGLPHNT 180
 DB 121 AKCVRENCIRYISILTOQAVYRDVAASVRKRPFTFNHLVLTAGLMTKEMAEFESIPSPHAK 180

QY 181 FVWPVPWVPANLSMKAYLGRIRDTVLQSLMNEVCTLRQCQGLYAYDWISIPLVY 240
 DB 181 YQWPMHLSMITLARDEGMISDIIYVDLMEKGRQFRVNLISLTFDWPVPLVYTVV 240

QY 241 TVAVYSFFLACIGROFLNPNKDYPGHENDLVVVPVFTILOFLYMGWLKVAEQLIN 296
 DB 241 HLAVRSYFLIALFGROYLHPESNRLNDFK-QTIDLVYVIMSLQLQFFIGMKVAVELN 299

QY 297 PFGEDDDDFETNWIIDRNLOVSLLSVDGMHONLPMERDMYNEAAP-QPPYTAASARR 356
 DB 300 PFGEDDDDFECNWLIDRNLOVSLLSVDGMHONLPMERDMYNEAAP-QPPYTAASARR 359

QY 357 HSPMGSTFNISLKKED-----LELMSKEADTD-----KESGYSS 392
 DB 360 NQVGSQCADMPTEEBEPFVPRPRRLSLRSMHWDMDTDVVPVGLKHTRDNSNYAS 417

RESULT 7
 ID O17205 PRELIMINARY; PRT; 557 AA.
 AC O17205;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE C01B12.5 protein.
 GN C01B12.5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R., Smaldon N., Smith A., Sonnenhammer E., Staden K., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.; "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans"; Nature 368:32-38(1994).
RL	Nature 368:32-38(1994).
RN	(2)
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RD	Scheet P., Maggi L.;
RE	"The sequence of C. elegans cosmid C01B12.";
RF	Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RG	[3]
RH	SEQUENCE FROM N.M.
RI	STRAIN=BRISTOL N2;
RJ	CA Waterston R.;
RK	RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RL	EMBL; AF025458; AAB70978.1; -
RM	InterPro: IPR000615; Worm_fam_8.
RN	Pfam; PF01062; DUF289; 1.
RO	PRODOM; PD002802; Worm.fam.8; 1.
RQ	SEQUENCE 557 AA; 64076 MW; 8A68BBB1436D0331 CRC64;
RS	Query Match 26.2%; Score 771.5; DB 5; Length 557;
RT	Best Local Similarity 31.4%; Pred.No.4.1e-58;
RU	Matches 17; Conservative 104; Mismatches 192; Indels 77; Gaps 9;
RV	
RY	
ZZ	
QY	1 MTITYTNKANVARLGSPFSSLLLRGSGTYSIKLLYGFLVFIFLYTSIRGLYRWLVSSDOOL 60
DB	1 MTVAYSILDVATSGLFTQIKVLLRWKGSVKWSINSELLIWLCCYSTLSVIYILLNKQAQE 60
QY	61 LPFKALAYCDSYIQILPISFVLGFVTUUVGRWMSQYENLPWRDLMTQVSFVGKDDEE 120
DB	61 VFEOQLCTFFDTFSFIPVTFMFGFVSIVYNRWTKVFDNVGMIDTSAITIAQYIRGTSEK 120
QY	121 GRLLRRTRLIRVAIGQGVILIRSISTSVYKRPTTLHLHLVLACFMTHGEHKQLQKLGLPHNT 180
DB	121 ARLIIRNCVRVYIMVAQWVFRDVSPAIRRRPTTKHLIGAGLLTEDELTEDAITSPQSK 180
QY	181 FWFPWWFWANLSMKAYLGGRIIRDTVLLQSLNNEVCTLRFCQGQLYAYDWISIPLYVTQVV 240
DB	181 YWOPIQMFLSVTVAKDBGLIADSVLYVDLIDKMRERFETKIILNVDFDMVIPLYVTQVV 240
QY	241 TVAVYSFELACLIGQFLNPKNKYPG--HEMDLVVPVFTTILOFIPLYMGWLKVASQLINPF 298
DB	241 NLAVRTYFVLALFGQFLNNNNIPGAWKWDIYFPIMTSLQIVFIVGLVKVSEVMNLPL 300
QY	299 GEODDDPETNWIIIDRNQLQVSLLSVDGHONLPMPMERDWMYNEAAPPPYTAASARSRRHS 358
DB	301 GEDDEDETNWIIERNLQGVYAVVDQAYGRYPFLKRDPDFWEDETPTQTDWDTSTTRKAHSH 360
QY	359 FMGSTFNWISKEDLELWSKEAADTKKESGY----SSTIGCFGLGQPKNYHLFLDKLKT 414
DB	361 MOGSCINWN-----EADLDNGLISYVRRRSRSGF-----DDDASSPSTENTRT 403
QY	415 KLICSKNPLEGQCKDANKQKQDWKFGLDLFKCVPRP-----KRGSGCGQPAPSS 468
DB	404 QSNASMLP-----RHMWQOOPRNKISAIAKPYKSAEPFRASVCVAAALNLN 449
QY	469 HPTEQAGAPSSSDTGDPSTDYOEICHMKKKTVFNLNIPESPTEHLQQRLRDQDMSTIOA 528
DB	450 HRR-----PSTDIREI-----EPIEN-----TDDVVINVPA 475
QY	529 LMKE 532
DB	476 LVQE 479


```

Db 939 RVFDFNFAALFQDTLNYILFTMLGPFVTVIIVGRWMDIFLNLGWVNDTALLIATYIRGSD 998
Qy 119 BEGRLLRRLTLYAILGVLILRSISTSVYKRPPTLHLVLAGFWTHGKHQLOKLG-L-P 177
Db 999 EKSRIMRRTALRYLVLTQVLFIRDLSLVYKRPPTLETIVVAGSFMLESEREYKAALSLGK 1058
Qy 178 HNTFVVPVWVFWANLKMAYLGRIRDTVLLQSLMNEVCTLRQCQGLYAYDWISIPLVYT 237
Db 1059 YPKYFLPQWCFSLYDARAQKIGADVWMLNLIKSVGDFRGLGOLLNDFWPIPLVYP 1118
Qy 238 QVTVVAVVSFFLACILGRQL-----NPNKDPYGHMDLVVP-VFTILOFLFYMGMLKVA 291
Db 1119 QVFLAVRIYFPLCLILARQSLVDIGKPKNSVPVYP---VPFLMTSLQFVYVGMKVA 1175
Qy 292 EQLINPFGEDDDDFETNWIIDRNLOVLSLVGDMQNLPPMRDMYWNAAPOPPYTAAS 351
Db 1176 ESLMPLGEDDDDFECNVLDRNLAVGLAIVDS-SEAAPHYKDVFWGLAIADPLYSDT 1234
Qy 352 ARSRHRSFWGSTFNISLKKEDLELWKEADTKKESGVSYSTIGC-----FLGLQPKN 404
Db 1235 ANIHINPOGSA--ATYETQNEIIMRPHVDINDMEDGDDVEGCNPROLSRFVSVSVN 1292
Qy 405 YHLPKDKTKLCKSKNPL 424
Db 1293 RNC-----ESRTSLSSRNPTI 1309

```

RESULT 11

```

Q9VRW4 ID Q9VRW4 PRELIMINARY; PRT; 731 AA.
AC AC Eukaryota; Metazoa; Arthropoda; Insecta;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE CG10173 protein.
GN CG10173
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1
RC SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.M.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berland B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Fogler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003563; AAF50668.1; -.
DR FlyBase; FBgn0035696; CG10173.
DR InterPro; IPR000815; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 731 AA; 80654 MW; FF4D625538826C6D CRC64;

Query Match 24.3%; Score 714.5; DB 5; Length 731;
Best Local Similarity 46.7%; Fred. No. 5e-53;
Matches 142; Conservative 50; Mismatches 103; Indels 9; Gaps 4;

Qy 78 ISFVLGFTVTVLVSRWWSQYENLPPMDRLMIQVSSFVEG-----KDEGRLLRRLTIRYA 132
Db 1 MSFVLGFTVTVLVSRWWSQYENLPPMDRLMIQVSSFVEG-----KDEGRLLRRLTIRYA 60

Qy 133 ILGQVLILRSISTSVYKRPPTLHLVLAGFWTHGKHQLOKLG--LPNHTFVVPVWVFWAN 190
Db 61 VLAVYITLQIRSLKVRFPPTQHLVDAGLMHSEMKIFEALNOKSPMSKYWMLVWATN 120

Qy 191 LSMKAYLGRIRDTVLLQSLMNEVCTLRQCQGLYAYDWISIPLVYTQVTVVAVVSFFLA 250
Db 121 IINRARKDGLIASDHIVQTLVLSLDIRRLGGLGYDTCVPLVYTQVTVVAVVSFFLA 180

Qy 251 CLIGRQFLNPKDYPGHE-MDLVVPVFTILOFLFYMGMLKVAEQLINPFGEDDDDFETNW 309
Db 181 ALLGRQMLPNVLDGRSDPDLFPFLFTVLQFVYVGMKVAEVLINPFGEDDDDFETNW 240

Qy 310 IIDRNLOVLSLVGDMQNLPPMRDMYWNAAPOPPYTAASARSRHRSFWGSTFNISL 368
Db 241 LIDRHIKAAYMIVDEMHHEHPPELLRDQWECVVPKDLFTYVASEHYRDKSPGSAEKYKV 300

Qy 369 KKED 372
Db 301 KKED 304

RESULT 12
Q9NA59 ID Q9NA59 PRELIMINARY; PRT; 551 AA.
AC Q9NA59;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Y73F8A.11 protein.
GN Y73F8A.11
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 1
RC SEQUENCE FROM N.A.
RX Matthews L.;
RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
DR EMBL; AL132862; CAB60536.1; -.
DR InterPro; IPR000615; Worm_fam_8.

```

DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 SQ SEQUENCE 551 AA; 64140 MW; 529FPREE63CB154EB CRC64;

Query Match 21.6%; Score 637; DB 5; Length 551;
 Best Local Similarity 28.3%; Pred. No. 1.7e-46;
 Matches 151; Conservative 79; Mismatches 136; Indels 168; Gaps 10;

Qy 1 MTITYTNKAVANARLGSSLLLCWRGSIYKLLYGELVFIFLYYSIRGLYRMVLSSDOOL 60
 Db 21 MTVNYGEVSSVFTFVKLLMTWRGSIKWSVRWELTWVAFVQVQVRYLMTEDQOK 80
 Qy 61 LFEKALYCDYSYI-QLIPISFVLGFTVTLVSRWWSQYENLFW- 102
 Db 81 FFEYSATHNLVRLTHIPLTFMLGFFVTIVDRMSVFTNIGFIENKYNVLKQENFYLIF 140
 Qy 103 -----PRLMT-- 108
 Db 141 QLSAVSYFFLKNNDPLTFNKNKTCNRLRYNTRSPIIQKITFTKTLTLLQEDKVVVLN 200
 Qy 109 ---QVSSFVEGKDEGRLLRRTLYAILGOVLILRSISTSVYKRFPTLHHLVLGAFMTH 165
 Db 201 VALSVGLTAVGTDHAAKVLRTIIRYLVLSQVLRLDSMRVRRFPPTMESLVSGLFLYR 260
 Qy 166 GEHKQLQKGLPHNTFWPVPWVFANLSMKAYLGGRIIRDVTLLQSLMNEVC----- 215
 Db 261 DELERMYKCDTMYNKYWLPTHMANQLVHKAMFETKNVDSV--QS-MNSVLMPIFSYLTVF 317
 Qy 216 -----TURTQ 220
 Db 318 FNFFLIIEVNSVPKLTNQEBAQALAEIYOKEDTIQFFYAFYLYGYGSIKEFRQS 377
 Qy 221 CGOLYADWISIPLYTQVTVVAVYSFELACLGROFLNPNKDYPGHEMDLWVPVETILQ 280
 Db 378 MEMTKYDWFIPIAYQVWPLAVRVYFIICLVSRQYL---LSAPPTAQASIVPIMTILQ 434
 Qy 281 FLFYMGWLKVAEQLINFFGDDDDFETNWIIDRLNQLVSLSDVGMHQNLPPMERD----- 335
 Db 435 FVPFVGWVKVAEALLNPLGDDDDFECNWLIDRNMTGIEIVDTCHDSCPLKLDPPDE 494
 Qy 336 ---MYW-----NEAAPQPYTAASRSR-----HSP-MGSTFNISLKED 372
 Db 495 SGTMYWCHPTKNGSQAPPYSAVIAISQHKNPFTKVLNVMVHPFSGTSSNSFROD 548

RESULT 13
 O62095 PRELIMINARY; PRT; 512 AA.
 AC O62095;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE C37A5-1 protein.
 GN C37A5-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; Z92828; CAB07337.1; -
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.

SQ SEQUENCE 512 AA; 58380 MW; A49PF8175CE1F0DCE CRC64;

Query Match 20.9%; Score 616; DB 5; Length 512;
 Best Local Similarity 28.3%; Pred. No. 1e-44;
 Matches 161; Conservative 103; Mismatches 200; Indels 104; Gaps 18;

Qy 1 MTITYTNKAVANARLGSSLLLCWRGSIYKLLYGELVFIFLYYSIRGLYRMVLSSDOOL 60
 Db 1 MTVSYNLDSVSPFFNFKLLFRWGSVWKSINSELVLLGLLVVYVYVALVTTEQA 60
 Qy 61 LFEKALYCDYSYI-QLIPISFVLGFTVTLVSRWWSQYENLFWPDRMLQVSSFVEGKDE 119
 Db 61 GVRKYIDHLQNDLQCKVPLTFMLAPFTIIVDRKNNFANIGFTENTAIATLVKGTG 120
 Qy 120 EGRLLRRTLYAILGOVLILRSISTSVYKRFPTLHHLVLGAFMTHGHEKQLQKGLPHN 179
 Db 121 DVLLAKTIRYLVLTQVLFVRDISLKVRRRFPNHDAILKAGFLQ--DHESIILGGDNGR 178
 Qy 180 T-FWVPWVFANLSMKAYLGGRIIRDVTLLQSLMNEVCVTLRTQCQLYAYDWISIPLVYTQ 238
 Db 179 TNYWMPVNWSSAILQKLFEDGNIPAAFLFNSVMQEVKTFRSNMATLCNYDWVPIAYPQ 238
 Qy 239 VVTAVVYSFELACLGROFLNPNKDYPGHEMDLWVPVETILQFLFYMGWLKVAEQLINPF 298
 Db 239 WFFAVRVYFPTCLFTFRQHLDMEDT---KTIDYFPILTIVFQTFPFMGWVKVAEALLNPL 295
 Qy 299 GEDDDDFETNWIIDRLNQLVSLSDVGMHQNLPPMERDYMNEAAPQPYTAASARS-RRH 357
 Db 296 GEDDDDFECNWLIDRNMTGMAIVNSKISDVPEMLADKF-NDPSIAPYYPEKVIDSGADH 354
 Qy 358 SPMGSTFNISLKEDLELWSKEADTDKQSGYSSTIGCFLGLQPKYHPLKDLKTKLL 417
 Db 355 ALVGSQAQTVL-----ASPNDI---IDMKVDL- 379
 Qy 418 CSKNPFLGEGCKDANQKQKQKQKGLDFLCKVPRFKRG---SHGQPOAPS-SHPTEQ 473
 Db 380 --NSPIVVGR---RNTNNTTIR------RLSSAFGRSHSVQHLGPEKPEPTSPFSQ 427
 Qy 474 S-AP-----SSSDTGDGSPSTDYQIECHMKKKKTV 501
 Db 428 SMAPQRYGAFELSNGLSGITQSQSHLPKLESEETD-PLPAQPVFVLPRLISE 486
 Qy 502 FNLNIPESPT-----EHLQORRLDQM 522
 Db 487 ----EPTSPTLPEDQTLFHVDDNNGLNKM 510

RESULT 14
 Q965X4 PRELIMINARY; PRT; 551 AA.
 ID Q965X4
 AC Q965X4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein Y37E11AR.1.
 GN Y37E11AR.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BRISTOL N2;
 RL MEDLINE=99069613; PubMed=9851916;
 RX None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium."
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Miller N., Maggi L.;
 RT "The sequence of C. elegans cosmid Y37E11AR."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 (3)

RN SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RL "Direct Submission.;"

RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC024759; AAK68431.1; -

DR InterPro: IPR000615; Worm_fam_8.

DR Pfam: PF01062; DUF289; 1.

DR ProDom: PD002802; Worm_fam_8; 1.

SQ SEQUENCE 551 AA; 63476 MW; 8B65140751455F25 CRC64;

Query Match

Best Local Similarity 19.0%; Score 559; DB 5; Length 551;

Matches 147; Conservative 91; Mismatches 168; Indels 110; Gaps 14;

QY 1 MTITYNKVANARLGSFSSLLLCWRGSIYKLYGELVFIFLYYSIRGLYRMVLSDDQOL 60

Db 97 MTVSNSDVSSTNFTLLRLRWKGSINXSVSELMMWCFYFLIALLYHKVLPND-- 154

QY 61 LPEKALVCDSYIQLIPISFVLGFTLVVSRWWSQYENLPWDRMIQVSSFVEGKDEE 120

Db 155 -----YYDS-----ALTNALIVSKDE 172

QY 121 GLLRRTLRVAILGOVLIRISITSVYKRPFTLHLVLGFWTHGEHKQLQKLGHPHT 180

Db 173 AKLIIRNIRYLVLAQILTFRDISIRVRRFPNIDSIKKAGFLTDNEELLEGIDLAYNK 232

QY 181 FWPVWVFANLSMKAYLGGRIRDTVLLOSLMNEVCTLRQCGLVAYDWISIPLVVTVV 240

Db 233 YWVPINWALSNOANSKGYVISAQWVSLIQEIKTFRNLGATICNDFCPIAYPQV 292

QY 241 TVAVSFFLACIGRQFLN-PNKVPGHEMDLVVPVFTILOFLYNGMLKVAEQLINPFG 299

Db 293 FFAVRIYELCLITROYIRVPNKELEPVQM-FIRPITIIIEICIVGWMKVAEALLNPLG 351

QY 300 EDDDDPEFNWIIDRLQVLSLSDVGHQHLPMERDMYNEAAPPYTAASARRH-S 358

Db 352 EDDDDFESNFDLIDKNIFTGMRIVD-QFDEAPFLFEDTP-SDPCAMPIYSEESQKNWQGA 409

QY 359 FMGSTFNLISLKKEDLELWKEADTDKESGYSTIGCFGLQPKNYHLPLKDLATKLLC 418

Db 410 LVGSVSNVTLAQFDN-----ITWVPVAPRLSIGDI----- 440

QY 419 SKNPLEGQCKDANKOKNDVWKPKGLDPLKCVPRFKRGSHCGPQAPS--SHPTQSA 476

Db 441 -----NAHDRSSIRR-----RFGSSQNCSEANSFREPVKVLRQP 476

QY 477 SSSDTGCPSTDYQEIChMKKKTVEFNINIPSPTE 512

Db 477 SDGSSNDAYELDYETVI-----KSVD-DL-VASSPTK 506

RESULT 15

017674

ID 017674 PRELIMINARY; PRT; 400 AA.

AC 017674; 046026;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE C49A1.2 protein.

GN C49A1.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Matthews L.;

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.;"

RL Nature 368:32-38(1994).

RN [3]

RP SEQUENCE FROM N.A.

RA Kershaw J.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z83221; CAB05710.1; -

DR EMBL; Z82095; CAB05710.1; JOINED.

DR EMBL; Z82095; CAB05029.1; -

DR EMBL; Z83221; CAB05029.1; JOINED.

DR InterPro: IPR000615; Worm_fam_8.

DR Pfam: PF01062; DUF289; 1.

DR ProDom: PD002802; Worm_fam_8; 1.

SQ SEQUENCE 400 AA; 46827 MW; D1CDF601E25E2D01 CRC64;

Query Match 17.1%; Score 503; DB 5; Length 400;

Best Local Similarity 35.6%; Pred. No. 4.2e-35;

Matches 122; Conservative 63; Mismatches 124; Indels 34; Gaps 10;

QY 1 MTITYNKVANARLGSFSSLLLCWRGSIYKLYGELVFIFLYYSIRGLYRMVLS---S 56

Db 1 MTITYRAVSTESISNFKLIIRHWSLAKSVQBEFLFWLPYVVCVYRCVLPIGWD 60

QY 57 DQQLF-EKALYCDSYIQLIPISFVLGFTLVVSRWWSQYENLPWDRMIQVSSFVE 115

Db 61 DQKKFVELSTHODMY---IPLEFMLGFTVTVDWRKSKQNIPIESCFAVSAALP 117

QY 116 GKDBEGRL-----LRRTLRVAILGOVLIRISITSVYKRPFTLHLVLGFM 163

Db 118 G---HGALEVSAGMDTKLTARRTIIRVLVSQILLFREFSVKVKGRFVDRSLVDSKEL 174

QY 164 THGEHKQLQK--LGLPHNTFWVPVWV-FANLSMKAYLGGRIRDTVLLOSLMNEVCTLRQ 220

Db 175 TENELTELSEELKTKRYDSYILPINWAFSILATE-----KLSNPQFMANVINDWQVK 229

QY 221 CGQLYAVDWISIPLVTVQVTVVAVYSFFLACIGRQFLNPNKDYPGHEMDLVVPVFTILQ 280

Db 230 LTLRNGDFIPIPLAYPQAVFLAVRFVLVCLFTRQHLDSL-DH--HAIDYFFPLITSFQ 286

QY 281 FLEYMGWLKVAEQLINPFGEDDDDEFTNWIIDRLQVLSLSDV 323

Db 287 FIFVGMKVAEILLNPMGEDDDDEFTNWIIDRLQVLSLSDV 329

Search completed: July 10, 2003, 12:26:22

Job time : 33.163 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 10, 2003, 12:21:29 ; Search time 32.0242 Seconds
(without alignments)
3763.954 Million cell updates/sec

Title: US-09-622-964-3
Perfect score: 3120
Sequence: 1 MTITYTSQVANARLGSFRL.....TKDHMPYALENDEAHS 585

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.1*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	54.5	428	Q8WNR7	Q8WNR7 sus scrofa
2	1137.5	36.5	465	Q8VCM0	Q8VCM0 mus musculus
3	1133.5	36.3	466	Q9NXP0	Q9NXP0 homo sapien
4	1009	32.3	721	Q9V3J6	Q9V3J6 drosophila
5	886	28.4	535	Q9VUM7	Q9VUM7 drosophila
6	814	26.1	612	Q17206	Q17206 caenorhabdi
7	770.5	24.7	557	Q17205	Q17205 caenorhabdi
8	758	24.3	584	Q18303	Q18303 caenorhabdi
9	742	23.8	327	Q9VUM6	Q9VUM6 drosophila
10	717.5	23.0	1447	Q16779	Q16779 caenorhabdi
11	694.5	22.3	731	Q9VRW4	Q9VRW4 drosophila
12	623.5	20.0	551	Q9NA59	Q9NA59 caenorhabdi
13	616.5	19.8	512	Q62095	Q62095 caenorhabdi
14	554	17.8	551	Q965X4	Q965X4 caenorhabdi
15	495	15.9	400	Q17674	Q17674 caenorhabdi
16	455	14.6	444	Q17671	Q17671 caenorhabdi

17	175	5.6	72	4	Q9BR80	Q9BR80 homo sapien
18	159.5	5.1	98	5	O01649	O01649 drosophila
19	117	3.8	516	16	Q82P99	Q82P99 salmonella
20	115	3.7	516	16	Q82772	Q82772 salmonella
21	113.5	3.6	700	10	Q94874	Q94874 oryza sativ
22	112	3.6	710	10	Q948J6	Q948J6 oryza sativ
23	112	3.6	1256	5	Q9W238	Q9W238 drosophila
24	111.5	3.6	798	4	Q9BWG7	Q9BWG7 homo sapien
25	109	3.5	1051	10	Q9S897	Q9S897 arabidopsis
26	107.5	3.4	1222	5	Q9TXV2	Q9TXV2 caenorhabdi
27	107	3.4	2606	12	O36414	O36414 alcelaphine
28	106.5	3.4	416	2	Q9F2A1	Q9F2A1 chromatium
29	105	3.4	683	11	Q99MT4	Q99MT4 mus musculu
30	105	3.4	786	11	Q99MT3	Q99MT3 mus musculu
31	105	3.4	1707	5	Q17652	Q17652 caenorhabdi
32	104.5	3.3	694	5	Q8T4D2	Q8T4D2 drosophila
33	104.5	3.3	785	10	Q9SA37	Q9SA37 arabidopsis
34	104	3.3	485	3	O13920	O13920 schizosacch
35	104	3.3	758	10	Q8S9K3	Q8S9K3 arabidopsis
36	104	3.3	886	3	Q9P7N3	Q9P7N3 schizosacch
37	103.5	3.3	356	4	Q9NUU6	Q9NUU6 homo sapien
38	103.5	3.3	741	16	Q8X7V5	Q8X7V5 escherichia
39	103	3.3	342	12	Q90387	Q90387 kaposi's sa
40	103	3.3	681	10	Q9SX10	Q9SX10 oryza sativ
41	103	3.3	758	10	Q9FN73	Q9FN73 arabidopsis
42	103	3.3	1359	5	Q8T6H8	Q8T6H8 dictyosteli
43	102.5	3.3	491	5	Q960V9	Q960V9 drosophila
44	102.5	3.3	569	5	Q9VL56	Q9VL56 drosophila
45	102.5	3.3	1002	3	O13849	O13849 schizosacch

ALIGNMENTS

RESULT 1

Q8WNR7 ID Q8WNR7 PRELIMINARY; PRT; 428 AA.
AC Q8WNR7; ID Q8WNR7; TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Bestrophin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_taxID=9823;
RN [1] SEQUENCE FROM N.A.
RP Marmorstein L.Y., McLaughlin P.J., Stanton B., Yan L., Crabb J.W.,
RA Marmorstein A.D.;
RT "Bestrophin interacts with the b-catalytic subunit of protein
phosphatase 2A".
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY064707; AAL40882.1; -
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
FT NON TER 1
SQ SEQUENCE 428 AA; 48498 MW; 3B5E91B2BCDB6383 CRC64;

Query Match 54.5%; Score 1701; DB 6; Length 428;
Best Local Similarity 74.8%; Pred. No. 7.4e-135;
Matches 324; Conservative 29; Mismatches 72; Indels 8; Gaps 4;

Qy	155	QHLVQAGFMPAEPKLEKLSLPHNMFWVPWFWANLSMKAWLGRIRDPILLQSLNEM	214
Db	2	QHLVQAGFMPSEHKLEKLSLPHNSFWMPWFWANLSMKAWLGRIRDPILLQSLNEM	61
Qy	215	NLTQTQCHLYAYDWISIPLVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHDLVVP	274
Db	62	NLTQTQCHLYAYDWISIPLVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHDLVVP	121
Qy	275	VFTFLQFFVYVGNLKVAEQLINPFGEDEDDDDPFTNWIVDRLQVSLLAVDHMDLPRMEP	334

```

Db 122 LFTFLQPPFYAGWLVAEQINLPFGEDDDDFETNIVDRSLQVSLAVDEMHDLPMPER 181
Qy 335 DWYWNKPQPPYTAASQAFPRASFMGSTFNISLNKEEMEFQNOEDDAHAGIIGREL 394
Db 182 DWYNDPPEHPYTAASQSRPSPFGSTFNISLKGDEMFOP--EEEEATGILGHP 239
Qy 395 GLOSHDHPHPRANSRTKLLWPRESLLHGLPKNHKAQKONVRGOEDNKAWKLKAVDAPK 454
Db 240 GLOSDHQPRTNSKTKLLWPKEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEG 296
Qy 455 SCPLYQRYGYVASAPOTPLSPMPFPLEPSAKLSHVTGID--TKOKSLKTVSSGAKS 512
Db 297 SAALCGRPGYHSAPOTPLGHTPMVFPPEBSAPLGRVRSGIDEAAKQSLQPATPSIKS 356
Qy 513 FELLSESCALMEHPEVSVORRKTVEFNLDMPETPEHNLKEPQSPNTNIHTLKDHMD 572
Db 357 FELLPESEASAEPLQSGHVRKTVEFNLDLSEAPE-HLKEPNLEPPNGIHALKDHMD 415
Qy 573 PYWALENDRDEAHS 585
Db 416 PYWALENDRDEAHS 428

RESULT 2
Q8VCM0 PRELIMINARY; PRT; 465 AA.
AC Q8VCM0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ20132.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019528; AAH19528.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 465 AA; 51998 MW; E7145E2F893B03F6 CRC64;

Query Match 36.5%; Score 1137.5; DB 11; Length 465;
Best Local Similarity 51.9%; Pred. No. 2.5e-87;
Matches 242; Conservative 56; Mismatches 127; Indels 41; Gaps 10;

Qy 50 YRLALTEEQQLMFEKLTLYCDSYIOLIPISFVLGFYVTLVTRWNNQYENLPWDLMSL 109
Db 7 YRFLAEEQRYFEKLVYCDQYASLIPVSFVLGFYVTLVTRWNNQYENLPWDLMSL 66
Qy 110 VSGFVEGKDEQSLRLRRLTIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHK 169
Db 67 VAGTVHGRDDRGLRYRRLTMRVAGLSAVLILRSVSTAVYKRPPTDHHVVEAGFMTREERK 126
Qy 170 QLEKLSLPHNMFVWVWFANLSMKAWLGGRIIDPILLOSLLNEMNTLTQCGHLYAYDW 229
Db 127 KFNENLSSYNKYWVCFWFSNLAQAARREGIRDSNLSALKLLLEELNVFRKCGMLPHYDW 186
Qy 230 ISPIVYTVQVTVAVYSPFLTCLVGRQFLNPAKAYPGHGLDLVVPVFTLQFFFYVGLWK 289
Db 187 ISPIVYTVQVTVAVSYFLACLIGRQFLDPAQGYKDHLDLCVPIFTLLQFFFYVGLWK 246
Qy 290 VAEQLINPFGEDDDDFETNIVDRNLQVSLAVDEMHDLPMPEDMYNKNKEPQPPYTA 349
Db 247 VAEQLINPFGEDDDDFETNIDRNPQVSMVAVDEMYYDLAVLEKDYWDAEAPAPYTA 306
Qy 350 ASA-QFRRASFMGSTFNISLNKEEMEFQ 376
Db 307 ATVEQLRQSPFGSTFDITLAKEDMQFQ 334

RESULT 4
Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

```

```

Db 307 ATAFLLQSPFSQSGTDFIALAKEDMQFQ--RLDQVGDPLGEVHGDFLQELL-----P 356
Qy 404 PRANSRTKLLPKRESLLHGLPKN-----HKAQKONVRGOEDNKAWKLKAVDAFKSGP 457
Db 357 AGAGSVGFL--GRRLLSLRR-----KNSCVSEASTASCAGAADGGVVECGGDDPLLDPS 411
Qy 458 LVQ-----RGYVASAPOTPL-SPTMPFF-----PLEPSAPSKL 489
Db 412 LREPELEPPACPEPPAPIPGTPPTTPTTIVSIPGRAPAPMPLPSPI 457

RESULT 3
Q9NXP0 PRELIMINARY; PRT; 466 AA.
AC Q9NXP0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CDNA FLJ20132 f18, clone COL06441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000139; BAA90970.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 466 AA; 52008 MW; BB9608363A09C2EC CRC64;

Query Match 36.3%; Score 1133.5; DB 4; Length 466;
Best Local Similarity 63.7%; Pred. No. 5.4e-87;
Matches 209; Conservative 50; Mismatches 68; Indels 1; Gaps 1;

Qy 50 YRLALTEEQQLMFEKLTLYCDSYIOLIPISFVLGFYVTLVTRWNNQYENLPWDLMSL 109
Db 7 YRFLTEGQRYFEKLVYCDQYASLIPVSFVLGFYVTLVTRWNNQYENLPWDLMSL 66
Qy 110 VSGFVEGKDEQSLRLRRLTIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHK 169
Db 67 VAGTVHGRDDRGLRYRRLTMRVAGLSAVLILRSVSTAVYKRPPTDHHVVEAGFMTREERK 126
Qy 170 QLEKLSLPHNMFVWVWFANLSMKAWLGGRIIDPILLOSLLNEMNTLTQCGHLYAYDW 229
Db 127 KFNENLSSYNKYWVCFWFSNLAQAARREGIRDSNLSALKLLLEELNVFRKCGMLPHYDW 186
Qy 230 ISPIVYTVQVTVAVYSPFLTCLVGRQFLNPAKAYPGHGLDLVVPVFTLQFFFYVGLWK 289
Db 187 ISVPLVYTVQVTVAVSYFLACLIGRQFLDPAQGYKDHLDLCVPIFTLLQFFFYVGLWK 246
Qy 290 VAEQLINPFGEDDDDFETNIVDRNLQVSLAVDEMHDLPMPEDMYNKNKEPQPPYTA 349
Db 247 VAEQLINPFGEDDDDFETNIDRNPQVSMVAVDEMYYDLAVLEKDYWDAEAPAPYTA 306
Qy 350 ASA-QFRRASFMGSTFNISLNKEEMEFQ 376
Db 307 ATVEQLRQSPFGSTFDITLAKEDMQFQ 334

RESULT 4
Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

```


Fri Jul 11 11:45:23 2003

01-JUN-2002 (Tremblrel. 21, Last annotation update)
 CG6264 protein (LP05915P).
 BEST OR CG6264.
 GN Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananatiades P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA de Pablo B., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA Cherry J.M., Cavley S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Fesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Golek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Taveanli B.C., Pappu K., Mardon G.;
 RT "deest" a Drosophila Homolog of Human Bestrophin.";
 RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragass V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF003686; AAF54503.1; -
 DR EMBL: AF218817; AAF32327.1; -
 DR EMBL: AY061546; AAL29094.1; -
 DR FlyBase; FBgn0040238; best.
 DR InterPro; IPR000615; Worm fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm fam_8; 1.
 SQ SEQUENCE 721 AA, 79642 MW, B37CD96DE2F15B15 CRC64;

Query Match 32.3%; Score 1009; DB 5; Length 721;
 Best Local Similarity 36.7%; Pred. No. 3.2e-76;
 Matches 250; Conservative 83; Mismatches 216; Indels 132; Gaps 18;
 QY 1 MTITYTQVANAR-LGSFSLRLCWRGSIYKLYGEEFLIFLLCYIIRFYRLALTBEQ 59
 DB 1 MTITYTGEVATCRGFCFLKLLLRWGSYKLVWLDLAFLLIYAINVMYVRFGLNPAQ 60
 QY 60 LMPEKLYCDSYIQLIPISFVLGFFYTVLVTRWNOYENLPWDRMLSLVSGFEGKDE 119
 DB 61 ETEPAIVQYCDYRELIPLSFVLGFFYTVLVTRWNOYTSIPWPDPIAFVSSNVHGD 120
 QY 120 QSRLLRRLRYANLGNVLILRSYSTAVYKFPSPQAHLVQAGTWPABHKOELK--SLP 177
 DB 121 RGRMRTTIRYVCLTVMLANVSPRVKFPGLNMLVEAGLNDNEKTIETMNAKFP 180
 QY 178 H-NMFWPVPWFANLKMKGIRIRDPILLIQLSLNEMNTLRTQGHLYAYDWISIPLY 236
 DB 181 RPSKHLPIVWASIIITRKEGRIRDDPAVKTIIDELNFKRGCGCLLSYDTISVPLV 240
 QY 237 TQVTVAVYSFPLTCLVGRQ-----FLNPAKAYPGHELDLVVPTFLQFFVY 285
 DB 241 TQVTVAVYSFPLTCLVGRQ-----KVDLYPFVFTTLOFFVY 292
 QY 286 GNLKVAEQLINPFGEDDDDDPETNIVDRNLOVSLAVDEMHDLPMEBDMYWNKPEP-Q 344
 DB 293 GNLKVAESLINPFGEDDDDDFEVNMVDRNLQVSLYLVDEMHHDPHLLKQDWVFPNE 352
 QY 345 PPYTAASAPRRASFMGSGTFNISLNK-----EEMEFQPNQDEEDAH----- 386
 DB 353 LPYTTAERFRENPEPSTAKIEVPKNAAMSTWSSVVIDEMADDASGHFSGAGNKML 412
 QY 387 -----AGIIGRFLG-----LQSHDHPPRANSRTKLLWPK 416
 DB 413 DSSPLSVSVGTLRVNTVSAALKKFLSRDSDRSGSATPSQDPYKFPASASLSGAV 472
 QY 417 RESLHGLPKNKAQKQVGOENKA--WKLKAVD-----AFKSGGLY 459
 DB 473 VGSATSAGKAPAGSLRITQCVIEVEDEQATITSMRANPRPNVMDIFATSSGAGTGLQ 532
 QY 460 QRPGYVSAQPNLSP-----TPMFFP-----LEPSAPS-----KLHVSVTG-- 494
 DB 533 PPAHSEPVDPISRPSPYNRAQSVPEPLPPGGVDALLSTAPAGSGPLLSNATAPS 592
 QY 495 --IDTKDKSLKTVSSGAKK---SPELLESB---GALMEHPVSVQVRKTVFNLTDMPE 546
 DB 593 SPVGSSEKSLYDPQKGSRETVESMDLSRSTLLGDAAVQPEDEG-----DDFDK 642
 QY 547 IPENHLKEPLOSPTNIHTTL 567
 DB 643 LKAEREKEKLMFQKNLARTI 663
 RESULT 5
 QYVUM7
 ID Q9VUM7 PRELIMINARY; PRT; 535 AA.
 AC Q9VUM7;
 DT 01-WAY-2000 (Tremblrel. 13, Created)
 DT 01-WAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE CG12327 protein.
 GN CG12327.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananatiades P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Ananatiades P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

Db 63 KLIFKWSGLKAIYLDIVMFCYAFISVIVRYALDRSQDPTFRFMQFCNRRLLDYPI 122
 Qy 79 SFVLGFYVTVTRWNOVENLWPDRLMSLVSGFVEGKDEOSRLRTRTLRIYANLGNVL 138
 Db 123 NFMLGFFVTVTRWNOVENLWPDRLMSLVSGFVEGKDEOSRLRTRTLRIYANLGNVL 182
 Qy 139 ILRSVSTAVYKRFPSAHLVQAGFWTPAHLKLEKLSLPHNNFWPWFVNFANLMSKAWLG 198
 Db 183 VFRDIHIGVRKRFPTLETWAAAGIMTSSELKYNVESRYAKYKWLGFNFNTLLNEARRE 242
 Qy 199 GRIRDPILLOSLNEMNTLRQCGHLYAYDMISILPVTVTVTVVAVYFFLTCVLNGROPL 258
 Db 243 GRISAYTQNAIAEIRTFRSLGLSIWTVPIPLMTYPOLVFMALHCYLYLCLVSROFV 302
 Qy 259 NPAXAYPGHELDLVVVFVTFQFPFYVGVKVAEOLINPFGEDEDDDFETNWTVDNLOVS 318
 Db 303 INSDAVNTTEIDLGVPFWTIEFIFYMGWLKVMADLLNPFGEDEDDDFCNFLIDNLTVA 362
 Qy 319 LLAVDENHODLPRMPEMDYNNKPEPQPPYTAASAQFRASFMGSTFN----ISLNKEEME 374
 Db 363 MGIVDDSHDDGPILKDMFNW-DTVSPLYSSAAQRNVNPFYGSATNADAQIPDDVRQIT 421
 Qy 375 FQPNQDEEDAHAGIIGFLGLOSHDHPHPRANSRTKLLPKRESLLHGLEPLKXKAAKQ 434
 Db 422 MIPHLNKEK-----LDQMYGKRT--NRPP-VQSVEL---KRDQRFSTG----- 459
 Qy 435 NVRGQEDNKAKLKAVIDAPKSGPLQYRPGYVSAPOTPLSPTMPFPLPSAPSAPKLSHVTG 494
 Db 460 -----NNRKOTLEAKFNKLGAMQKXHSKSLTTFPESST-----KARASTD 501
 Qy 495 IDTKDKSLKTVSSGAKSKFELLSESDGALMEHPVSQVRRKTVFNLNLTMPPEIPENHLKE 554
 Db 502 IET---VSTPENTQKPCY-----SNPDCEIVEDEESQ-----DIPKVTPEPKQE 542
 Qy 555 P 555
 Db 543 P 543
 RESULT 9
 Q9VUM6 PRELIMINARY; PRT; 327 AA.
 ID Q9VUM6
 AC Q9VUM6;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE CG7259 protein.
 GN CG7259.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S.E., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.M., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Staden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 DR EMBL; AB03531; AAF9649.1; --
 DR FlyBase; FGN0036491; CG7259.
 DR InterPro; IPR00615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 SQ SEQUENCE 327 AA; 38594 MW; BE7CE3BC05370F46 CRC64;
 Query Match 23.8%; Score 742; DB 5; Length 327;
 Best Local Similarity 46.3%; Pred. No. 3.1e-54;
 Matches 142; Conservative 55; Mismatches 92; Indels 18; Gaps 4;
 Qy 68 YCDSVLIQIPISFVLGFGYVTVTVTRWNOVENLWPDRLMSLVSGFVEGKDEOSRLRRT 127
 Db 3 YCEKNGSLIPISFVLGFGYVTVTVTRWNOVENLWPDRLMSLVSGFVEGKDEOSRLRRT 62
 Qy 128 LIRYANLGNLIRSVSTAVYKRFPSAHLVQAGFWTPAHLKLEKLSLPHNNFWPWFVNFANLMSKAWLG 182
 Db 63 ILRYVCLQVIVTFWISPRVKERFPTTYQIIEAGFLENEKKIIEAMDQAFSPYKHI--W 120
 Qy 183 VFWVWFANLMSKAWLGGRIRDPILLOSLNEMNTLRQCGHLYAYDMISILPVTVTVTV 242
 Db 121 MPVWAAISIVMRARRENKIRDDYAVKTIIDELNQLRNGCGFLLYYDWSVPLVTVTVTV 180
 Qy 243 AVYSFELCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFYVGVMLKVAEQLINPFGEDE 302
 Db 181 ATYSFELCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFYVGVMLKVAEQLINPFGEDE 240
 Qy 303 DDFETNWTVDNLOSLNEMNTLRQCGHLYAYDMISILPVTVTVTVTVTVTVTVTVTVTV 351
 Db 241 DDFELNWTVDNLOSLNEMNTLRQCGHLYAYDMISILPVTVTVTVTVTVTVTVTVTVTV 300
 Qy 352 AQFRAS 358
 Db 301 AYMESQS 307
 RESULT 10
 O16779 PRELIMINARY; PRT; 1447 AA.
 ID O16779
 AC O16779; O16781;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Hypothetical 164.2 Kda protein.
 GN T21D12.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.

RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of <i>Drosophila melanogaster</i> ";
RL	Science 287:2185-2195(2000).
DR	EMBL; AE003563; AAF50668.1; -
DR	FlyBase; FBgn0035696; CG10173.
DR	InterPro; IPR00615; Worm_fam_8.
DR	Pfam; PF01062; DUF289; 1.
DR	ProDom; PD002802; Worm_fam_8; 1.
SQ	SEQUENCE 731 AA; 80654 MW; FF4DG25528826C6D CRC64;

Query Match	
Best Local Similarity 23.3%; Score 694.5; DB 5; Length 731;	
Matches 174; Conservative 80; Mismatches 195; Indels 67; Gaps 12;	

Qy	78	ISFVLGYPTVTVTRWNQYNLWPPDRMLSLVSFGVEG-----KDEQSRLRLRTLRYA	132
Db	1	MSFVLGYTNLVKKRWEEQYRLLPWPDTLALFSAATPSNSGGVNNETGLWRNINRYM	60
		: : : : :	
Qy	133	NLGNVLTLSRVSTAVYKFPESAHLVQAGFMTPAEHKLEKLS--LPHNFWPWVWFAN	190
Db	61	VLAYVIITQRISLRVKGRFPTTQHLVDAGLMHSEMKIFEALNQSPSKYMWPLVWATN	120
		: : : : :	
Qy	191	LSMKAWLGGRIIRDPIILQSLNLNMTLRTCCHLYAYDWISIPLVYTQVTVVAVYSFLT	250
Db	121	IINRKADGLIASDHIVOTILVELUSDIRRRUGLIGDYTCVPLVYTQVTVTLVITYFIA	180
		: : : : :	
Qy	251	CLVGROQLNPAKAYPGHE-LDLVVVFTEFLQFFFYVWGMLKVAEQLINPFGEDDDDPFTNW	309
Db	181	ALLGRQMLPNVLDRSRGREDPDLPFLFTVLQFVYVWGMLKVAEVLINPFGEDDDDIELNW	240
		: : : : :	
Qy	310	IVRNQLQSVLLAVDEMHDQILPRMEPDMYWNKPQP-PYTAASAQFRASFGMGTFNISL	368
Db	241	LIDRHIIKAAVMIVDMEHZEHPPELLRDQYECVPKDLPYTVASEHYARKDEPKGSAEKYV	300
		: : : : :	
Qy	369	NKEEMEQQNPQEDEDHAHGI--CRFGLQSHDHPFRANSRTKLLMPKRESLLHGLEPK	427
Db	301	KGEDA-----MYAINMPGGGKRMLSDDDIYADYESVDTEMVRERRKNWLVRQLSR	349
		: : : : :	
Qy	428	NHKAANKVNRQGQEDNKAWLKAVDAFKSGPL-YORPGYSAPOPTPLSPTPMFPFLEP---	483
Db	350	MGSNRSOST-----AYSGGMPFNRLNLSVYSPESGSLPTLILOOQOLQ	394
		: : : : :	
Qy	484	-----SAPSKLHSVTGITDKDKLKTVSSGAKKSFEILLSDSGDALMEHPVSQVRR	534
Db	395	QAHOQQAGSQPSK-SSLYGKFVHRKSIARAQQLIKONSKLNGLVNVVAKTRPRI-----	448
		: : : : :	
Qy	535	KTVFNLTMDMPETPENHLKBPLOS-----PTNIHTT	566
Db	449	-----PTPEVAKDGNTPATSSVLMAPQOOLSTT	476
		: : : : :	

```

RESULT 12
Q9NA59
ID Q9NA59 PRELIMINARY; PRT; 551 AA.
AC Q9NA59;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE W73F8A.11 protein.
OS Y73F8A.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RN RP SEQUENCE FROM N.A..
RN RA Matthews L.;
RN RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN RP SEQUENCE FROM N.A..

```

RX	MEDLINE=99069613; PubMed=9851916;
RA	none;
RT	"Genome sequence of the nematode C.elegans: A platform for investigating biology,"
RL	Science 282:2012-2018(1998).
DR	EMBL; ALJ32862; CAB60536.1; --
DR	InferPro; IPR000615; Worm_fam_8.
DR	Pfam; PF01062; DUF289; 1.
DR	ProDom; PD002802; Worm_fam_8; 1.
SQ	SEQUENCE 551 AA; 64140 MW; 529PFEEG3CB154EB CRC64;

Query Match 20.0%; Score 622.5; DB 5; Length 551;
Best Local Similarity 28.3%; Pred. No. 7.8e-44;
Matches 151; Conservative Mismatches 147; Indels 159; Gaps 11;

Qy	1	MTITYTSQVANRILGSFRIILLCWGRSGIYKLLYGEPFLIFLCYYIIIRFIYRALALTEEQOL	60
Dd	21	MTVNYNQEVSSNSFTPVKLMTWRGSIWKSVRWELTMWIAFAVVQMVRYLMTEDQQK	80
Qy	61	MFEKLTLYCDSYQLIPISFVLGPYTVLVTRW-----WQVEN-----	99
Dd	81	FPEYSAILHLNVLRIHIPLTLMGLFFVTIVDVRMRSVFNTIGFIENKTNLVKQENFYLI	140
Qy	100	-----LPWPDLMSL- 109	
Dd	141	QLSNVSYPFLKNNDFLTNTNKTCRNLRYNRTSPILIOKITFKTKILLQEEOKVVVLN	200
Qy	110	-----VSGFVEGDEQSRLLRTRIIRYANLGNVILRSVSTAVYKRFPFSAQHLYQAQMTP	165
Dd	201	VALSVGTLVAGTDHAAKVLARTIIRYLVLSQLVLRDISMRVRRRPFTMESLVSGGFYLR	260
Qy	166	AEHKOLEKLSLPHNMFWVPWWFANLSMKAWLGGRIIDPI-----	205
Dd	261	DELERYTKDTNKKYLLPHTHANQLVHKAFETKNVDVSQSMSNVLMPISYLTVPFNF	320
Qy	206	-----LLOSLINEMNTL-----RTOCGH 223	
Dd	321	FLIIEVNVSVPKLTNOEAEQATALAEIYQKEDIQFFYAFTLYGYGSNIKEFRQSMEM	380
Qy	224	LXYADWLISPLVTVQVTVNAVYSEFLTCLVGRQFLNPAPAYPGHELDLVVPVTFLOPPF	283
Dd	381	LTKYDWMPIDPIAPVQVYFLAVRVYFICLVSRQYLLSA---PPTAQSIVPIIMTILOFVF	437
Qy	284	YVGWLKVAEQLINPFGEDDDDFETNWITVDRLNQLVSLIADVMEHQDLPRM---EPD----	335
Dd	438	FVGMKVAEALLNPLGEDDDDFECNMLIDRMNSTGTIEVTDCHDCSPCLKLDEFDRSGET	497
Qy	336	MYWNKP-----EPOPPYTAASAQ-----FRRASFNGSFTNISIANKEMEFAQ----PN	378
Dd	498	MYWCHPTKNGSQAPPASAVAIISOHNKFTKVLNMVHPFSSTGNSNSSSFRODPNN	551

RESULT	ID	OG2095	PRELIMINARY;	PRT;	512 AA.
OG2095	AC	OG2095;			
DT	DT	01-AUG-1998	(T-EMBLrel. 07, Created)		
DT	DT	01-AUG-1998	(T-EMBLrel. 07, Last sequence update)		
DT	DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)		
DE	DE	C37A5.1	protein.		
GN	GN	C37A5.1.			
OS	OS	Caenorhabditis elegans.			
OC	OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	OX	NCBI_TaxID=6239;			
RN	RN	[1]			
RP	RP	SEQUENCE FROM N.A.			
RA	RA	White S.;			
RL	RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.			
RN	RN	[2]			
RP	RP	SEQUENCE FROM N.A.			
RX	RX	MEDLINE=99069613; PubMed=9851916;			
RA	RA	none.			

```

RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z92828; CAB07337.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 512 AA; 58380 MW; A49F8175CE1F0DCE CRC64;

Query Match
Best Local Similarity 19.8%; Score 616.5; DB 5; Length 512;
Matches 152; Conservative 94; Mismatches 217; Indels 53; Gaps 13;

QY 1 MTITVTSOVANARLGSFSLLCWRSIYKLYGFLFLCYIIRFYRLALTEEQOL 60
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
1 MTSVSNLDSVSSVFFNPKLLFRWRGWSKSIWSELVLLGLLYLVWYRVALTTEQKA 60

QY 61 MPEKLTLYCDSYI-OLIPISFVLGFVTVLVTRWNQYENLPWPDRLMSLVSGFVGKDE 119
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
61 GVRKYIDHLQNDLCKVPTFMALFVTIIVDRKMFANIGFIENATIAIATLVKGTGEG 120

QY 120 OSRLRLRTIIRYANLGNVILRSVSTAVYKRPFSQHLVQAGFMTPAEHKQLEKLSL--- 176
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
121 DVLLAKRTIIRYLVLTQVLVPRDISLKVRRRFPNDHAIKAGFL--QDH---ESIILGCD 175

QY 177 -PHNFWPWTWFWANLSKAWLGGRIIRDPILOSLLNEMNTLRTQCGHLYAYDWTISIPLV 235
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
176 NGRTNYMFWNVSSAILQKLFQDGNIPAPLFSNVQEVKTPRSNNATLCNDYDWPPIA 235

QY 236 YQVTVAVYSFPLTCLVGRQFLNPAKAVPGHELDLVVPVTFLOPFFVVGWGLKVAEQLI 295
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
236 YQVTVAVYSFPLTCLVGRQFLNPAKAVPGHELDLVVPVTFLOPFFVVGWGLKVAEQLI 295

QY 296 NPFGEDDDDFENWIVDRNLQVSLAVDEMHDQLPRMPEDMYWNKPEPPY----- 347
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
293 NPLGEDDDDFECNLYIDRNATGMAIVNSKYSQVDEPLADKF-NDPSYAPYPERKVIDSG 351

QY 348 -----TAASAQFRASFMGSTFNISLNKEMEFQPNQOEDEDAHAGIIRGFLGLOSH- 399
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
352 ADHALVGAQVTLAEPNDIIDMMKVDLNSPIVVGRRNTNTSTIRRLSSAF-GRRSHS 410

QY 400 -DHHPPRANSRTKLMPKRESLLHGLPKNKAQKQNVGRQ-EDNKAWKLKAVDAFKSG- 456
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
411 VOHLGPEKPEPT-----PFSQMAPQRYGAFELNGFNLSGLSGITTSQS 456

QY 457 --PLYQRPGYSA-----QTPLSPTMPFFLEPSAPS 487
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
457 HLPKLESESETDPLAQQPVFLPRTLSEPTSTPT 492

RESULT 14
Q965X4 PRELIMINARY; PRT; 551 AA.
AC Q965X4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein Y37ELLAR.1.
GN Y37ELLAR.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;
RA Miller N., Maggi L.;
RT "The sequence of C. elegans cosmid Y37ELLAR.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024759; AAK68431.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 551 AA; 63476 MW; 8B65140751455F25 CRC64;

Query Match
Best Local Similarity 17.8%; Score 554; DB 5; Length 551;
Matches 146; Conservative 91; Mismatches 190; Indels 102; Gaps 13;

QY 1 MTITVTSOVANARLGSFSLLCWRSIYKLYGFLFLCYIIRFYRLALTEEQOL 60
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
97 MTSVSNLDSVSSVSTNFRLLRLRWKSIWKSSELMMWCFYFLIALLYHKVLPEND-- 154

QY 61 MPEKLTLYCDSYIQLIPISFVLGFVTVLVTRWNQYENLPWPDRLMSLVSGFVGKDEQ 120
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
155 -----YDSAL-----TVNAIIVSKCDE 172

QY 121 SLLRLRTIIRYANLGNVILRSVSTAVYKRPFSQHLVQAGFMTPAEHKQLEKLSLPHNM 180
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
173 ARLIRNRIIRYLVLAQILTPDISIRVRRFPNIDSIKKAGFLTNEEBELLEGIDLAYNK 232

QY 181 FWPMWFWANLSKAWLGGRIIRDPILOSLLNEMNTLRTQCGHLYAYDWTISIPLVTVQV 240
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
233 YWVFINWALSUNQANSKGYISAPGWSLQIETKTPRNGLATCNFQWCPIPIAYPOV 292

QY 241 TVAVYSFPLTCLVGRQFLNPAKAVPGHELD-----LVVPVTFLOPFFVVGWGLKVAEQLIN 296
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
293 FFAVRIYFLCLITRQYIR----VFNKELEPVQMFIRPFIIFICIVGMKVAEALIN 348

QY 297 PFGEDDDDFENWIVDRNLQVSLAVDEMHDQLPRMPEDMYWNKPEPPYPTAASQ-FR 355
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
349 PLGEDDDDFENFLIDKNIFTGMRIYVQDFEAPLFEDT--FSDPGAMPIYSEESQKNYQ 406

QY 356 RASFMGSTFNISLNKEMEFQPNQOEDEDAHAGIIRF-LG-LOSHDHHPPRANSRTKL 413
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
407 NGALVGSVSNVTL-----AQFDENITWVPVAPRLSISGDIINAH-----RSSIRRRFK 453

QY 414 WPKRESLLHGLPKNKAQKQNVGRQEDNKAWL-----KAVDAFKSGPLYQRPGYSA 467
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
454 SGSSQNCSTRANSFREPKNVLRQPSDGSNDAYELDYETVKSVDLVAS----- 502

QY 468 PQTPLSPTMPFFLEPSAPSALH-SVTGIDTKSKLTKTVSSGAKKSFEL 515
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
503 -----SPTKPGKFTSLTKVDENDEIPVSLIFAKKSLGL 537

RESULT 15
O17674 PRELIMINARY; PRT; 400 AA.
ID O17674;
AC O17674; O46026;
DT 01-NOV-1998 (TrEMBLrel. 05, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE C49A1.2 protein.
GN C49A1.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;

```


Search completed: July 10, 2003, 12:26:17
Job time : 36.0242 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:19:13 ; Search time 26.8587 Seconds
(without alignments)
2158.111 Million cell updates/sec

Title: US-09-622-964-5
Perfect score: 2350
Sequence: 1 MTITYSQVANRLGFSRL.....QLPPSSVEPLWAPPSTSA 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2350	100.0	435	20 AAY29954	Human CGICE short
2	1535	65.3	585	20 AAY29953	Human CGICE long f
3	1297	55.2	551	20 AAY29955	Mouse CGICE protei
4	763.5	32.5	721	22 ABB62213	Drosophila melanog
5	685	29.1	535	22 ABB63121	Drosophila melanog
6	543.5	23.1	327	22 ABB63135	Drosophila melanog
7	493	21.0	731	22 ABB64914	Drosophila melanog
8	450	19.1	368	22 ABB68609	Novel human digno
9	361	15.4	314	22 AAB70085	Human secreted pro
10	338	14.4	701	22 ABB08608	Novel human digno

11	239	10.2	251	22 AAB70067	Human secreted pro
12	239	10.2	251	23 ABB65506	Human albumin fusi
13	238	10.1	99	22 AAU22652	Novel human colon
14	238	10.1	99	22 AAM92658	Human digestive sy
15	204.5	8.7	377	22 ABB66411	Novel human digno
16	174	7.4	30	22 AAY29956	Human CGICE partia
17	151	6.4	118	22 ABB04059	Human musculooskele
18	144	6.1	108	22 AAM90422	Human immune/haema
19	141.5	6.0	72	22 AAO13117	Human polypeptide
20	138	5.9	76	22 AAB43272	Human ORFX ORF3036
21	137	5.8	60	22 ABB03675	Human musculooskele
22	136.5	5.8	78	22 AAM86740	Human immune/haema
23	136.5	5.8	102	22 AAB95008	Human protein sequ
24	135	5.7	71	23 ABB79293	Human prostate spe
25	135	5.7	93	22 AAO13000	Human polypeptide
26	134	5.7	43	22 AAM85912	Human immune/haema
27	134	5.7	65	22 AAG75572	Human colon cancer
28	133	5.7	278	23 ABB51443	Human MDDR SEQ ID
29	132	5.6	76	21 AAB45219	Human secreted pro
30	131.5	5.6	90	22 AAU30514	Novel human secret
31	130.5	5.6	63	21 AAB63223	Gene 43 human secr
32	130.5	5.6	68	21 AAB44753	Gene 23 human secr
33	130.5	5.6	71	21 AAB44759	Gene 50 human secr
34	130.5	5.6	72	21 AAB45218	Gene 42 human secr
35	130.5	5.6	74	21 AAB51464	Human secreted pro
36	130.5	5.6	74	21 AAB51466	Human secreted pro
37	130.5	5.6	80	22 AAO12799	Human polypeptide
38	130	5.5	40	22 AAG75544	Human colon cancer
39	130	5.5	76	22 AAG9856	Human prostate CDN
40	130	5.5	76	22 AAU69933	Human prostate CDN
41	130	5.5	76	22 AAM01211	Human prostate-spe
42	130	5.5	76	22 AAM01288	P789P splice varia
43	130	5.5	76	22 AAB95316	Human P776P splice
44	130	5.5	76	23 ABB95393	Human P776P splice
45	130	5.5	107	22 AAO12697	Human polypeptide

ALIGNMENTS

RESULT 1
AAY29954
ID AAY29954 standard; Protein: 435 AA.
XX
AC AAY29954;
XX
DT 22-NOV-1999 (first entry)
XX
DE Human CGICE short form protein sequence.
XX
KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;
KW BMD; age-related macular dystrophy.
XX
OS Homo sapiens.
XX
PN WO9943695-A1.
XX
PD 02-SEP-1999.
XX
PF 22-FEB-1999; 99WO-US03790.
XX
PR 25-FEB-1998; 98US-0075941.
PR 18-DEC-1998; 98US-0112926.
XX
PA (MERI) MERCK & CO INC.
XX (UYUP-) UNIV UPPSALA.
XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;
XX WPI; 1999-540560/45.
XX N-FSDB; AAZ21228.
XX
PT Human and mouse polynucleotides encoding CGICE polypeptides

XX PS Claim 7; Fig 5; 67pp; English.

XX CC The present sequence represents the human CGICE protein. When the CGICE

XX CC gene is mutated it is responsible for Best's macular dystrophy (BMD).

XX CC Polynucleotides encoding CGICE are useful for diagnosing whether a

XX CC patient carries a mutation in the CGICE gene. Normal and mutated

XX CC CGICE proteins are useful for identifying activators and/or inhibitors

XX CC of these proteins, in order to treat BMD. The CGICE gene offers a

XX CC simpler and cheaper method of diagnosing BMD without the need for the

XX CC presence of the patient. The gene may also be useful to discovering

XX CC the genetic cause of age-related macular dystrophy.

XX SQ Sequence 435 AA;

Query Match 100.0%; Score 2350; DB 20; Length 435;

Best Local Similarity 100.0%; Pred. No. 3.3e-237;

Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

DB 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

QY 61 MPEKLTLYCDSYIQLIPISFVLGPFYVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

DB 61 MPEKLTLYCDSYIQLIPISFVLGPFYVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

QY 121 GRLLRRTLIIRYANLGNVLIRSVTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

DB 121 GRLLRRTLIIRYANLGNVLIRSVTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

QY 181 FWVPWVFWANLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 240

DB 181 FWVPWVFWANLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 240

QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOPFYVGLKVSRLALLGWR 300

DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOPFYVGLKVSRLALLGWR 300

QY 301 HGQRHGQQLLETRMOCQERKVSRYESSQAWRTVPVTPATREAEAGESLEPGRRRLWQS 360

DB 301 HGQRHGQQLLETRMOCQERKVSRYESSQAWRTVPVTPATREAEAGESLEPGRRRLWQS 360

QY 361 SSSSTPLERMMLIRPTGLSTGICRCPCWLMWRCRTCLGWSRTCTGISPSHSPTQLPP 420

DB 361 SSSSTPLERMMLIRPTGLSTGICRCPCWLMWRCRTCLGWSRTCTGISPSHSPTQLPP 420

QY 421 SSVPEPLWAPPSTSA 435

DB 421 SSVPEPLWAPPSTSA 435

RESULT 2

AA29953

ID AAY29953 standard; Protein; 585 AA.

XX AC AAY29953;

XX AC AAY29953;

DT 22-NOV-1999 (first entry)

XX DE Human CGICE long form protein sequence.

XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

XX KW BMD; age-related macular dystrophy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 121

FT /note= "encoded by GGC"

XX PN WO9943695-A1.

XX XX

PD 02-SEP-1999.

XX XX

PF 22-FEB-1999; 99WO-US03790.

XX XX

PR 25-FEB-1998; 98US-0075941.

PR 18-DEC-1998; 98US-0112926.

XX XX

PA (MERI) MERCK & CO INC.

PA (UYUP-) UNIV UPPSALA.

XX XX

PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;

XX XX

DR WPI; 1999-540560/45.

DR DR N-PSDB; AA221227.

XX XX

XX Human and mouse polynucleotides encoding CGICE polypeptides -

PS Claim 7; Fig 3; 67pp; English.

XX XX

CC The present sequence represents the human CGICE protein. When the CGICE

CC gene is mutated it is responsible for Best's macular dystrophy (BMD).

CC Polynucleotides encoding CGICE are useful for diagnosing whether a

CC patient carries a mutation in the CGICE gene. Normal and mutated

CC CGICE proteins are useful for identifying activators and/or inhibitors

CC of these proteins, in order to treat BMD. The CGICE gene offers a

CC simpler and cheaper method of diagnosing BMD without the need for the

CC presence of the patient. The gene may also be useful to discovering

CC the genetic cause of age-related macular dystrophy.

XX SQ Sequence 585 AA;

Query Match 65.3%; Score 1535; DB 20; Length 585;

Best Local Similarity 99.7%; Pred. No. 1.2e-151;

Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

DB 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

QY 61 MPEKLTLYCDSYIQLIPISFVLGPFYVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

DB 61 MPEKLTLYCDSYIQLIPISFVLGPFYVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

QY 121 GRLLRRTLIIRYANLGNVLIRSVTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

DB 121 GRLLRRTLIIRYANLGNVLIRSVTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

QY 181 FWVPWVFWANLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 240

DB 181 FWVPWVFWANLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 240

QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOPFYVGLKVSRLALLGWR 290

DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLOPFYVGLKVSRLALLGWR 290

RESULT 3

AA29955

ID AAY29955 standard; Protein; 551 AA.

XX AC AAY29955;

XX AC AAY29955;

DT 22-NOV-1999 (first entry)

XX DE Mouse CGICE protein sequence.

XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

XX KW BMD; age-related macular dystrophy.

XX OS Mus sp.

XX PN WO9943695-A1.

XX XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL07224.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions

PS Disclosure; SEQ ID NO 16155; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 535 AA;

Query Match 29.1%; Score 685; DB 22; Length 535;
Best Local Similarity 42.9%; Pred. No. 1.1e-62;
Matches 126; Conservative 66; Mismatches 98; Indels 4; Gaps 2;

OY 1 MTITYTSQVAN-ARLGSFRLLLCMRGSYKLLYGEFLIFLLCYIIRIYRIALTEEQO 59

DB 1 MTVSYTAEVATCSHFQCFWKLWRWFASIKIIVDLLAFSLSCFYPMAYIYVALRDVXK 60

OY 60 LMFKEKLYCDYSYIQLIPISFVLGFFVTLVTRWNNQYENLWPDRLMSLVSGFVEGKDE 119

DB 61 PVFEDIVMYCHSYNLIPLSFVLGFFVYGIIBRWNNQYITVPDPPLAVVYSALVRGQDE 120

OY 120 QGRLLRRTLIRYANLGNVLILRSVSTAVYKRPFSQAHLVQAGFMTPAEHKQLEKLSL--- 176

DB 121 HGRLLRRTIMRVCLALTWLSMISPVIKRRPFTDQLLEVLGNANINMKAMDVKFP 180

OY 177 PHNMFVWPVWPFANISMKAWLGGRIIRDPILOSLLNEMNTLRQCCHLYAYDWISIPLVY 236

DB 181 KHPKYWMPVWAAISIVTRARKEGRWDWDFSLKSMIDELNKRAGCNMLHYDTISVPLVY 240

OY 237 TOVTVVAVYSPLTCLVGRQFNLPAKAYPGHELDLVVVFVTFPLQFFFFYVGLKV 290

DB 241 TOVTVLAVYSVFASIFGHQWIDRIDIKYNNIVSYFFPLFSTLEFFFFFGWGLKV 294

RESULT 6

ABB63135

ID ABB63135 standard; Protein; 327 AA.

XX ABB63135;

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 16197.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL07238.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions

PS Disclosure; SEQ ID NO 16197; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 327 AA;

Query Match 23.1%; Score 543.5; DB 22; Length 327;
Best Local Similarity 44.3%; Pred. No. 3.9e-48;
Matches 101; Conservative 48; Mismatches 72; Indels 7; Gaps 2;

OY 68 YCDSYIQLIPISFVLGFFVTLVTRWNNQYENLWPDRLMSLVSGFVEGKDEQGRLLRRT 127

DB 3 YCKNGSLIPLSFVLGFFVTRVWNRWDQYTTIPWPGIALISTSIHSGDDRAVMRT 62

OY 128 LIRYANLGNVLILRSVSTAVYKRPFSQAHLVQAGFMTPAEHKQLEKLSL-----SLPHNMF 182

DB 63 ILRVYVCLQVITFTWISPRVVRFPFTYQITIEAGFLENEKKIIEAMDQAFPSYPKH--W 120

OY 183 VPWFWFANLSMKAWLGGRIIRDPILOSLLNEMNTLRQCCHLYAYDWISIPLVYVTVV 242

DB 121 MPVWAAISIVTRARKEGRWDWDFSLKSMIDELNKRAGCNMLHYDTISVPLVY 180

OY 243 AVYFFFTCLVGRQFNLPAKAYPGHELDLVVVFVTFPLQFFFFYVGLKV 290

DB 181 ATYSFFFLPSVLGQWNSHSDGTRRRRWFPIITLVLFQFFFGWGLKV 228

RESULT 7

ABB64914

ID ABB64914 standard; Protein; 731 AA.

XX ABB64914;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 21534.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PP 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL09017.
 DR XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 21534; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL161175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 731 AA;
 SQ
 Query Match 21.0%; Score 493; DB 22; Length 731;
 Best Local Similarity 45.7%; Pred. No. 2.3e-42;
 Matches 101; Conservative 36; Mismatches 76; Indels 8; Gaps 3;
 QY 78 ISFVLGFVYTVVTRWQYENLPWDRMLSLVSGFVEG-----KDEGRLLRRLIRYA 132
 Db 1 MSFVLGFVYVNLVVRWWEQYRLLPWPDTLALFISAAIENSNGVNNETGRLMRRNINYM 60
 QY 133 NLGNVLILRSVSTAVYKRPFAQHLVQAGFTPAEHKOLEKLS--LPHNMFVWPMWPFAN 190
 Db 61 VLAVITLQRLSLVRKRPFTTQHLVDAGLHSEMKIFEALNQKSPSKYAMPLVWATN 120
 QY 191 LSMKAWNGRIRDPILQSLNEMNTLTQCHLYAYDWISIPLYTQVTVVAVYSFELT 250
 Db 121 IINRAKDGLIASDHIVQTLVELSDIRRLGLLIGYDVCVPLVYTVQVTVLVLYTYIA 180
 QY 251 CLVGRQFLNPAKAPGHE-LDLVVPVFTFLQFFVYVGLKV 290
 Db 181 ALLGRQMLPNVLDKSGREDPDLFPFLPTVLOFVYVGLKV 221
 RESULT 8
 ABG08609
 ID ABG08609 standard; Protein; 368 AA.
 AC ABG08609;
 XX 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #8600.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PP 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR N-PSDB; AAS72796.
 DR XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID NO 38968; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 368 AA;
 SQ
 Query Match 19.1%; Score 450; DB 22; Length 368;
 Best Local Similarity 44.6%; Pred. No. 2.8e-38;
 Matches 91; Conservative 27; Mismatches 66; Indels 20; Gaps 4;
 QY 162 FMTFAEHKOLEKLSLPHNMFVWPMWPFANLSMKAWNGRIRDPILQSLNEMNTLTQ 221
 Db 1 FMTTDERKLFNHLKSPHLKYWVPFIWFGNLATKARNEGRIKSDVLOSQSLMTMNRYSWC 60
 QY 222 GHLYADWISIPLYTQVTVVAVYSFELTCLVGRQFLNPAKAPGHELDLVVPVFTFLQF 281
 Db 61 SLFLGYDWVGIPLVYTVQVTVLAVYFFACLIQGFQDPTKGYAGHDLDLVIFITLQF 120
 QY 282 FFYVGLKVLGSLRALLGWRHQRGHGQQLLETRMQCKQKRVSRVSSQAWWRTVPVAPTR 341
 Db 121 FFYAGWLKV--AEQLI---NPFGEDDDDFTFN-CIDRLNQ-----VSLLA 160
 QY 342 EAEAGESLEPGRRLWQSSSTP 365
 Db 161 VDEHMSLPKMKKDIYWDDSAARP 184

RESULT 9
 AAB70085
 ID AAB70085 standard; Protein; 314 AA.
 XX
 AC AAB70085;
 XX
 DT 14-MAY-2001 (first entry)
 XX
 DE Human secreted protein-related polypeptide #6.
 XX
 KW Human; secreted protein; immunomodulatory; anti-sclerotic;
 KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
 KW immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
 KW antialzheimer's; antiparkinsonian; antimicrobial; vulnery; gene therapy;
 KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
 KW neurological; infection.
 XX
 OS Homo sapiens.
 XX
 PN WO200112776-A2.
 XX
 PD 22-FEB-2001.
 XX
 PF 15-AUG-2000; 2000WO-US22350.
 XX
 PR 16-AUG-1999; 99US-0148759.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;
 XX
 DR WPI; 2001-244245/25.
 XX
 PT Nucleic acids encoding 18 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy.
 XX
 PS Disclosure; Page 377-378; 380pp; English.
 XX
 CC The present sequence is given in a specification relating to nucleic
 CC acid molecules encoding 18 novel human secreted proteins. The nucleic
 CC acids and proteins may be used in the prevention, diagnosis and
 CC treatment of diseases including immune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. The
 CC nucleic acid molecules may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples. The polypeptides may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators of
 CC protein expression and activity.
 XX
 SQ Sequence 314 AA;
 Query Match 15.4%; Score 361; DB 22; Length 314;
 Best Local Similarity 43.6%; Pred. No. 4, 8e-29;
 Matches 88; Conservative 14; Mismatches 78; Indels 22; Gaps 4;
 QY 172 EKLSLPHNMFVWVWFWANLSKMWLGGRIRPILLOSLNENWTLRTGCHLYADWIS 231
 DB 13 ESLKSDFNKYVPCVWFTNLAQARDGRDRIIDIALCLLELNKYRACSKMLFHYDWIS 72
 QY 232 IPLVTVQVTVAVYGFLLTCLVGRQFLNP-----AKAY-----LDLVVPVF 276
 DB 73 IPLVTVQVTVAVYGFLLTCLVGRQFLNP-----AKAY-----LDLVVPVF 132

QY 277 TPLQFFFYVGLKVLGSLRALLGWRHQRGCGOOLLETRMQCQKYSRV-----ESSQ 329
 DB 133 TLLQFFFYVGLKVLGSLRALLGWRHQRGCGOOLLETRMQCQKYSRV-----ESSQ 192
 QY 330 AWRRTFVTPATREAEAGESLEP 351
 DB 193 YWDEDPQPPYTVATAESLRP 214

RESULT 10
 ABG08608
 ID ABG08608 standard; Protein; 701 AA.
 XX
 AC ABG08608;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #8599.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS72795.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID No 38967; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostic, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 701 AA;
 Query Match 14.4%; Score 339; DB 22; Length 701;
 Best Local Similarity 26.7%; Pred. No. 3, 7e-26;

20-OCT-2000; 2000US-0240960.
 20-OCT-2000; 2000US-0241221.
 20-OCT-2000; 2000US-0241785.
 20-OCT-2000; 2000US-0241786.
 20-OCT-2000; 2000US-0241787.
 20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-465567/50.
 N-PSDB; AAS39532.
 Isolated polypeptide for treating, preventing and/ or prognosing
 disorders related to the colon including colon cancers and also for
 testing and detection e.g. diagnosis -
 Claim 11; SEQ ID No 429; 562pp; English.
 The present invention relates to the isolation of novel human colon
 associated polypeptides, and the cDNA (AAS39348-AAS39581) and genomic

sequences encoding for them. The sequences of the invention are useful
 in the diagnosis, treatment, prevention and/or prognosis of disorders
 of the colon including colon cancer, congenital abnormalities
 (e.g. atresia and stenosis), bacterial and viral infections,
 inflammatory bowel disease (IBD), neoplastic cell disorders
 (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
 colonic inflammation, diarrhoea and dysentery, malabsorption syndromes
 (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
 The polynucleotide sequences of the invention can also be used in gene
 therapy. AAU22468-AAU22701 represent the novel human colon associated
 polypeptides of the invention.
 Note: the sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 99 AA;
 SQ
 Query Match 10.1%; Score 238; DB 22; Length 99;
 Best Local Similarity 55.4%; Pred. No. 7.5e-17;
 Matches 46; Conservative 14; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MTITVTSOVANARLGSFRLLLCWGSIYKLLYGEFLICLYIIRFYRLALTEEQOL 60
 Db 13 MIVSYTLKVAERAGFGSGLLLRWGSIYKLLYKFLFLFGALYAVLSITYRLLLTQEQRY 72
 QY 61 MFEKLTLYCDSYIQLIPISFVLG 83
 Db 73 VYAQVARYCNSADLIPLSFVLG 95
 RESULT 14
 AAM92658
 ID AAM92658 standard; Protein; 99 AA.
 XX
 AC AAM92658;
 XX
 DT 05-NOV-2001 (first entry)
 XX Human digestive system antigen SEQ ID NO: 2007.
 DE Human; digestive system antigen; gene therapy; cancer; appendicitis;
 KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
 KW digestive system disorder; Meckel's diverticulum.
 XX Homo sapiens.
 OS WO200155314-A2.
 FN 02-AUG-2001.
 XX
 PD 17-JAN-2001; 2001WO-US01324.
 XX
 PF 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.

QY 61 MFEKLTLYCDSYIOLIPISFVLG 83
 Db 73 VVAQVARYCNSADLIPLSFVLG 95
 RESULT 15
 ABG06411
 ID ABG06411 standard; Protein; 377 AA.
 XX AC ABG06411;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #6402.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX FN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI: 2001-639362/73.
 XX DR N-PSDB; AAS70598.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations,
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX PS Claim 20; SEQ ID No 36770; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 377 AA;

Query Match 8.7%; Score 204.5; DB 22; Length 377;
 Best Local Similarity 30.1%; Pred. No. 1.5e-12;
 Matches 65; Conservative 26; Mismatches 64; Indels 61; Gaps 8;

Search completed: July 10, 2003, 12:24:12
 Job time : 28.8587 secs

Db 51 MTVTYARVANARFGFSQQLLLWEGSIYKLLWRLLCFLGFYMAISAAAYRAGLQCAAE 110
 QY 57 EQLMFEKLTLYCDSYIOLIPISFVLGFIYVTLVTRWNNQYENLPWDRMLSLVSGFVVG 116
 Db 111 GKRYFEKLVIVCDH---MPASSLS-----PSCLLMLLSEKQS 146
 QY 117 KDEGRLRLRTLLRYANLG-----NVLLRSYSTAVYKRFPSAQHLVQAGF---MTPA 166
 Db 147 LAQENAGLRERMRPEGETPGLTAKKLLILQS-----QLEQLQENFRALVRPR 196
 QY 167 EHKQLEKLSLPH--NMFWVP-----VWVFA 189
 Db 197 PQALILQAEIHLVRKFWTPNSGSGPEAKVIYDINFA 232

QY 1 MTITYTSQVANARLGFSRLLLCWRGSIYKLLYGEFLIFLCYIIIRPIYR----LALTE 56

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:34 ; Search time 9.96817 Seconds
(without alignments)
1283.984 Million cell updates/sec

Title: US-09-622-964-5
Perfect score: 2350
Sequence: 1 MTITYTSQVANRLGFSRL.....QLLPSSVEPPLWAPPSTSA 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgm2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgm2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgm2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgm2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgm2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	5.5	76	4	US-09-605-785-575
2	111	4.7	52	4	US-09-227-357-537
3	97	4.1	681	2	US-08-272-255-6
4	97	4.1	681	2	US-08-272-255-9
5	97	4.1	681	5	PCT-US95-08565-6
6	97	4.1	681	5	PCT-US95-08565-9
7	95.5	4.1	582	3	US-08-906-865-3
8	95.5	4.1	582	4	US-09-129-668-3
9	95	4.0	342	3	US-08-785-928-1
10	95	4.0	342	3	US-08-728-603-17
11	94	4.0	681	3	US-08-964-268-5
12	92.5	3.9	353	1	US-08-229-418-2
13	92.5	3.9	353	2	US-08-932-761A-2
14	92.5	3.9	353	4	US-09-307-912-2
15	92.5	3.9	353	5	PCT-US95-04464-2
16	91.5	3.9	1692	4	US-09-263-933-4
17	91.5	3.9	2307	4	US-09-263-933-2
18	89.5	3.8	1692	4	US-09-263-933-18
19	89.5	3.8	2013	1	US-08-324-977-12
20	89.5	3.8	2013	2	US-08-384-616-12
21	89.5	3.8	2013	2	US-08-904-686A-12
22	89.5	3.8	2013	4	US-09-315-850-12
23	89.5	3.8	2307	4	US-09-263-933-16
24	89.5	3.8	2620	1	US-08-324-977-32
25	89.5	3.8	2620	2	US-08-384-616-32
26	89.5	3.8	2620	2	US-08-904-686A-32
27	89.5	3.8	2620	4	US-09-315-850-32

28	89.5	3.8	2621	1	US-08-324-977-36	Sequence 36, Appl
29	89.5	3.8	2621	2	US-08-384-616-36	Sequence 36, Appl
30	89.5	3.8	2621	2	US-08-904-686A-36	Sequence 36, Appl
31	89.5	3.8	2621	4	US-09-315-850-36	Sequence 2, Appl
32	89.5	3.8	3010	1	US-08-324-977-2	Sequence 2, Appl
33	89.5	3.8	3010	1	US-08-324-977-14	Sequence 14, Appl
34	89.5	3.8	3010	2	US-08-384-616-2	Sequence 2, Appl
35	89.5	3.8	3010	2	US-08-384-616-14	Sequence 14, Appl
36	89.5	3.8	3010	2	US-08-904-686A-2	Sequence 2, Appl
37	89.5	3.8	3010	2	US-08-904-686A-14	Sequence 14, Appl
38	89.5	3.8	3010	4	US-09-315-850-2	Sequence 2, Appl
39	89.5	3.8	3010	4	US-09-315-850-14	Sequence 14, Appl
40	88.5	3.8	1692	4	US-09-263-933-11	Sequence 11, Appl
41	88.5	3.8	2307	4	US-09-263-933-9	Sequence 9, Appl
42	86.5	3.7	3010	4	US-09-014-416-3	Sequence 3, Appl
43	86	3.7	509	3	US-08-845-546-2	Sequence 2, Appl
44	85	3.6	294	1	US-08-142-439A-4	Sequence 4, Appl
45	85	3.6	294	2	US-08-869-477-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-605-785-575
; Sequence 575, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-575

Query Match 5.5%; Score 130; DB 4; Length 76;
Best Local Similarity 63.2%; Pred. No. 4.1e-06;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

OY 318 QERKVSRYESSQAWRTVPVTPATREAEAGESLEPGRRR 355
Db 3 KSRFTKTKITQAWRPVPTREAEAGESLEPGRLR 40

RESULT 2
US-09-227-357-537
; Sequence 537, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.

[illegible]

Db 374 QLPWRGMKGFWDITLLD-----ADLESALGQVITGTLPLDSREFRIDNPQEG 423
QY 347 ESLEPGRR--RLWQSSSTPLERM-----MMILRPTGLSTG 381
Db 424 YKFDNGBYVRRWLPSELRLPTDWHHPWNPAPESVLQAAGIELG 467

RESULT 4

US-08-272-255-9
; Sequence 9, Application US/08272255
; Patent No. 5824859
; GENERAL INFORMATION:
; APPLICANT: Cashmore, Anthony R.
; APPLICANT: Ahmad, Margaret
; APPLICANT: Lin, Chentao
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
; TITLE OF INVENTION: Using the Same
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/272,255
; FILING DATE: 08-JUL-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: UPN-1795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 681 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-272-255-9

Query Match 4.1%; Score 97; DB 2; Length 681;
Best Local Similarity 21.5%; Pred. No. 0.23;
Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;

QY 123 LLRRTLYRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFMTPAEHKOLEKLS 175
Db 267 LVRIKQVAVANEGNEAGESVNLFLKSGISREYSRISFNH-----PYSH---ERPL 315
QY 176 LPHNMFVWPVWFANLSMKAWLGGRIIRDPILOSLLNEMNTLRTCCGHLAYDWTISIPLV 235
Db 316 LGHLKFF-PAVDENY-FKAWRQRTGYP-LVDAGMRE-----LWATGWLH-358
QY 236 YTVVTVVAVYSFFTLCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFVYVGLKVLGSLRA 295
Db 359 --DRIRVVVSFFVKVL-----373
QY 296 LLGWRHGQRGHGQQLLETRMOCQERKVSRESSQAWMR--TPVIPATREAE-----AG 346
Db 374 QLPWRGMKGFWDITLLD-----ADLESALGQVITGTLPLDSREFRIDNPQEG 423
QY 347 ESLEPGRR--RLWQSSSTPLERM-----MMILRPTGLSTG 381

Db 424 YKFDNGBYVRRWLPSELRLPTDWHHPWNPAPESVLQAAGIELG 467

RESULT 5

PCT-US95-08565-6
; Sequence 6, Application PC/TUS9508565
; GENERAL INFORMATION:
; APPLICANT: Cashmore, Anthony R.
; APPLICANT: Ahmad, Margaret
; APPLICANT: Lin, Chentao
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
; TITLE OF INVENTION: Using the Same
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08565
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/272,255
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: UPN-1795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 681 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-08565-6

Query Match 4.1%; Score 97; DB 5; Length 681;
Best Local Similarity 21.5%; Pred. No. 0.23;
Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;

QY 123 LLRRTLYRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFMTPAEHKOLEKLS 175
Db 267 LVRIKQVAVANEGNEAGESVNLFLKSGISREYSRISFNH-----PYSH---ERPL 315
QY 176 LPHNMFVWPVWFANLSMKAWLGGRIIRDPILOSLLNEMNTLRTCCGHLAYDWTISIPLV 235
Db 316 LGHLKFF-PAVDENY-FKAWRQRTGYP-LVDAGMRE-----LWATGWLH-358
QY 236 YTVVTVVAVYSFFTLCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFVYVGLKVLGSLRA 295
Db 359 --DRIRVVVSFFVKVL-----373
QY 296 LLGWRHGQRGHGQQLLETRMOCQERKVSRESSQAWMR--TPVIPATREAE-----AG 346
Db 374 QLPWRGMKGFWDITLLD-----ADLESALGQVITGTLPLDSREFRIDNPQEG 423
QY 347 ESLEPGRR--RLWQSSSTPLERM-----MMILRPTGLSTG 381
Db 424 YKFDNGBYVRRWLPSELRLPTDWHHPWNPAPESVLQAAGIELG 467

RESULT 6

PCT-US95-08565-9
 ; Sequence 9, Application PC/TUS9508565
 ; GENERAL INFORMATION:
 ; APPLICANT: Cashmore, Anthony R.
 ; APPLICANT: Ahmad, Margaret
 ; APPLICANT: Lin, Chien-ao
 ; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
 ; TITLE OF INVENTION: Using the Same
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
 ; STREET: One Liberty Place, 46th floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/08565
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/272,255
 ; FILING DATE: 08-JUL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Leary Ph.D., Kathryn
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: UPN-1795
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 568-3100
 ; TELEFAX: (215) 568-3439
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 681 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-08565-9

Query Match 4.1%; Score 97; DB 5; Length 681;
 Best Local Similarity 21.5%; Pred. No. 0.23;
 Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;
 QY 123 LLRRLIRYANIGNV-----LILRSVSTAVYKRPSPSAHLVQAGFMTPTAEHKOLEKLS 175
 Db 267 LVRIKQVAVANEENAGEESVNLFLKSLGRYSRIISFNH-----PYSH-----ERPL 315
 QY 176 LPHNMFVPPVWFWANLSKMAWLGGRIRDPILLOSLNEMNTLRTQGHLYAYDWTISIPLV 235
 Db 316 LGHLKFP-FWADENY-FKAWRQGRGTYP-LVDAGMRE-----LWATGWLH----- 358
 QY 236 YQVVTVAVYSEPLTCLVGRQFLNPAKAYPGHELDLVVFTFLOFFVVGWLVGLSRA 295
 Db 359 --DRIVVYVSFFVKV----- 373
 QY 296 LLGWRHGQRGQOQLLETRMQCQKRVSRVSSQAWWR--TPVIPATREAE-----AG 346
 Db 374 QLPWRGWMKYFDWTLDD-----ADLESALGQVITGTLDPDSREFRIDNPQEG 423
 QY 347 ESLPGR--RLWQSSSTPLERM-----MMILRPTGLSTG 381
 Db 424 YKFDPNGEYVRWLPELSRLPTDTHHPWNAPESVLQAAGIELG 467

RESULT 7
 US-08-906-865-3
 ; Sequence 3, Application US/08906865
 ; Patent No. 6040168

; GENERAL INFORMATION:
 ; APPLICANT: Greengard, Paul
 ; APPLICANT: Porton, Barbara
 ; APPLICANT: Kao, Hung-Teh
 ; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: David A. Jackson, Esq.
 ; STREET: 411 Hackensack Ave, Continental Plaza, 4th
 ; STREET: Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/906,865
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-202
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 582 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; DESCRIPTION: /desc = "Synapsin Ila"
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; US-08-906-865-3

Query Match 4.1%; Score 95.5; DB 3; Length 582;
 Best Local Similarity 20.8%; Pred. No. 0.26;
 Matches 85; Conservative 53; Mismatches 135; Indels 135; Gaps 19;
 QY 110 VSGFVGEKDEGRLLRRLIRYAN---LGNVLILRSVSTAVYKRPSPSAHLVQAGFMTPTA 166
 Db 135 VLGDYDIKVEQAEFSELNLVAHADGTVAVDMQVLN-GTKVRSFRPDFVLIRQHAFCMA 193
 QY 167 EHKOLEKL-----SLP-----HNMFWPPVWFWANLSKMAWLGGRIRDPILLOSL 211
 Db 194 ENEDFRHLIIGQVAGLPSINSLSIYFCDKPPVFAQLVAIYKTLGGE-KPFLIEQTY 252
 QY 212 ---NEMTLRT-----OCGHLYA-----YDWISIPLVYQVTVVAVYSPFL----- 249
 Db 253 PNHKEMLTLPFPVVKIGHAHSQKVKVNHVDPODIASVVALTQTYATAEFPIDSKY 312
 QY 250 -----TCLVGRQFLNPAKAYPGHELDLVV-----VFTFLQFFFYVG 286
 Db 313 DIRVQKIGNNYKAYMRTSISGNMKTNTGSAM-----LEQIAMSRYKLWVDTCEMF---G 365
 QY 287 WLKVLGSLRALLGWRHGGORH-----GQQLLETRMQCQKRVSRVSSQAWW 332
 Db 366 GLDCAVKV-----HGKDGKDYIFEVNDCSNPLIGEHQVEDRQITELVISKM--NQLLS 419
 QY 333 RTPVIP-----ATREAEAGESLEPGRRLRWQSSSTPLERMMILRPTGLSTGICRCP 387
 Db 420 RTPALSPQRPLTTQPPQSGTLKOP-----DSSKTPPQRP-----PPQG----- 457
 QY 388 WLWNRCTRTCLGWSRTCTGISPSHSPPTQLLPPSSVBPPLWAPPSTSA 435

Db 458 -----GPGQPGMPPGKVLPPRLPPGSLPSSSS 489

RESULT 8

US-09-129-668-3
; Sequence 3, Application US/09129668B
; Patent No. 6429010
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE AND USES
; FILE REFERENCE: 600-1-202 CIP
; CURRENT APPLICATION NUMBER: US/09/129,668B
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 08/906,865
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-129-668-3

Query Match 4.1%; Score 95.5; DB 4; Length 582;
Best Local Similarity 20.8%; Pred No. 0.26;
Matches 85; Conservative 53; Mismatches 135; Indels 135; Gaps 19;
QY 110 VSGFVEGKDEQGRLLRLIRVAN---LGNVLIRSVSTAVYKRPFSQHLVQAGFMTPA 166
Db 135 VLGDYDIKVEQAEFSELNLVAHADGTAYAVDMQVLEN-GTKVRSRPDPVLIRQAFGWA 193
QY 167 EHKQLEK-----SLP-----HMFVTPWTFANLSKMLGLGRIRDPILLOSL 211
Db 194 ENEDRHLIIGQYAGLSINSLSIYNFCDPKPVFAQVLAITYKTLGGE-KEPLIEQTVY 252
QY 212 ---NEMNTLRT-----QCCHLYA---YDWISIPLYVTQVTVAVYSFPL----- 249
Db 253 FNHKEMTLTPFPVVVKGHAHSGMGKVENHYDFQDIASVVALTQYVATAPEDISKY 312
QY 250 -----TCLVGRQFLNPAKAYPGHBLDLPV-----VFTFLOPPFFVYG 286
Db 313 DIRVQKIGNYKAVNRTSISGNWKTNTGSAM-----LEQIAMSRYKLWDTCTSEMF--G 365
QY 287 WLKGLSRALLGWRHGRGH-----CQQLLETMOQERKVSREVSQAWW 332
Db 366 GLDICAVKAV-----HGKDKDYIFVMDCSMPLGEHQVEDRQLITELVISKM--NOLLS 419
QY 333 RTPVIP-----ATREAEAGESLEPCRRRLWMOSSSTPLERMMLIRPTGLSTGICRCPC 387
Db 420 RTPALSPORLTTQPOSGTLKDP-----DSSKTPQRP-----PPQG----- 457
QY 388 WLWNRCTKCLGWSKTCGIFSGHSHPPTQLPPSVSEPLWAPPTSA 435
Db 458 -----GPGQPGMPPGKVLPPRLPPGSLPSSSS 489

RESULT 9

US-08-785-928-1
; Sequence 1, Application US/08785928
; Patent No. 6087115
; GENERAL INFORMATION:
; APPLICANT: Gershengorn, Marvin C.
; APPLICANT: Arvanitakis, Leandros
; APPLICANT: Geras-Raaka, Elizabeth
; APPLICANT: Cesarman, Ethel
; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE G PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE
; TITLE OF INVENTION: ANTAGONISTS OF G PROTEIN COUPLED RECEPTORS
; NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,928
FILING DATE: 22-JAN-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1320
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-785-928-1

Query Match 4.0%; Score 95; DB 3; Length 342;
Best Local Similarity 24.6%; Pred. No. 0.14;
Matches 69; Conservative 41; Mismatches 110; Indels 60; Gaps 15;
QY 3 ITTYSQVANARLGFSRLL--CWRGSIYKL-LYGEFLIFLLCYIIRFIYRLALTEEQ 59
Db 68 VTIYFKHRSRAGADIDILLGICLNSCLSLSLAEVLMFLFPNIISGLCEL-----E 121
QY 60 LMPEKLTLYCDSYQLDIPISFVLGVYTLVTRWNOYENLPWDRMLSLVSGFVEGKD- 118
Db 122 IFFYLYVYLDIP-SVVCVSLV-RYLLVAYSTRSNPKKQSLGWLTSAAALLIALVLSGDA 179
QY 119 --EGRCL--LRRTLIRYANLGNV-----LILRSYSTAVYKRPFSQHLVQAGFMTPAEH 168
Db 180 CRHSRVVDPSVKQAMCYENAGNMTADWRLHVRIVS-----VTAGFLLP----- 223
QY 169 KOLEKLSLPHMFWPVPWFWFANLSKMLGLGRIRDPILL-----QSLLNEMNTL----- 217
Db 224 --LALLILFYALTWCV-VRRTKLQARRKVRGVIVAVVLLFFVFCFPHYVNLNLLDILLRRR 280
QY 218 --RTQCCHLYADWISIPLVYT---QVTVAVYSFETCL 252
Db 281 WIRDSC---YTRGLINVLAVTSLLOALYSAAVPLIYSCL 317

RESULT 10

US-08-728-603-17
; Sequence 17, Application US/08728603
; Patent No. 6093806
; GENERAL INFORMATION:
; APPLICANT: Cesarman, Ethel
; APPLICANT: Knowles, Daniel M.
; TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA

STATE: New Jersey
 COUNTRY: USA
 ZIP: 07065-0907
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/307,912
 FILING DATE: 10-May-1999
 CLASSIFICATION: <UNKNOWN>
 ATTORNEY/AGENT INFORMATION:
 NAME: Hand, J. Mark
 REGISTRATION NUMBER: 36,545
 REFERENCE/DOCKET NUMBER: 19202CA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (732) 594-3905
 TELEFAX: (732) 594-4720
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 353 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-307-912-2

Query Match 3.9%; Score 92.5; DB 4; Length 353;
 Best Local Similarity 19.6%; Pred. No. 0.26;
 Matches 74; Conservative 45; Mismatches 108; Indels 151; Gaps 18;

QY	38	IFLLCYII---RFIYRLALTEBQQLMFEKLTLYCDSYIQLIPISFVLG--FYVTLVWTR	92
Db	44	IISICFFGLGNLFVLLVFLPRQLNVAEIVL---ANLAASDLVFLGLPFAENI---	97
QY	93	WNOYENLWLP-----DRMSLVSGFVEGKDEQGR	123
Db	98	WNOFN---WPFGLLCRVINGVIKANLFIISFLVVAISQDRYRVLHPMASGRQRRRQ	153
QY	124	LRRLIRYANLGNVILRSVAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHMFW	183
Db	154	ARVTCVLIVWVGGL---SIFTLRSIQVDPDLNITACI-----LLLPH----	195
QY	184	PWVFNLSMKAWLGGRIID-----PI-----LQSLNEMNTLTTCGHLYA	226
Db	196	-----EAMHPARIVELNIGLFLPLAAIVFFNVHILASLTRREVSRTCCG--R	243
QY	227	YDMISIPLVYTVVAVYSFPLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFVY-	285
Db	244	KDSKTTALITLVVA-----FLVCWAPYHF-----PAPLEFLFQVQ	279
QY	286	-----GWLKVGLSRA-----LLGWRHGQRGHGOOLLETRMOCQERK	321
Db	280	AVRGCFWEDFDLGLQLANFAFTNSLNPNVIYFVG-----RLPRTKWELYKQCTPKS	334
QY	322	VSREVS-----QAWNR	333
Db	335	LAPISSSHRRKEIFQLFWR	352

RESULT 15

PCT-US95-04464-2
 ; Sequence 2, Application PC/TUS9504464
 ; GENERAL INFORMATION:
 ; APPLICANT: Linemeyer, David L.
 ; APPLICANT: Menke, John G.
 ; APPLICANT: Hees, John P.
 ; APPLICANT: Borkowski, Joseph A.
 ; APPLICANT: Bierillo, Kathleen K.
 ; TITLE OF INVENTION: DNA ENCODING BRADYKININ B1 RECEPTOR
 ; NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
 ADDRESSEE: John W. Wallen III
 STREET: P.O. Box 2000
 CITY: Rahway
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07065
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04464
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Wallen III, John W.
 REGISTRATION NUMBER: 35,403
 REFERENCE/DOCKET NUMBER: 19202
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-3905
 TELEFAX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 353 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-04464-2

Query Match 3.9%; Score 92.5; DB 5; Length 353;
 Best Local Similarity 19.6%; Pred. No. 0.26;
 Matches 74; Conservative 45; Mismatches 108; Indels 151; Gaps 18;

QY	38	IFLLCYII---RFIYRLALTEBQQLMFEKLTLYCDSYIQLIPISFVLG--FYVTLVWTR	92
Db	44	IISICFFGLGNLFVLLVFLPRQLNVAEIVL---ANLAASDLVFLGLPFAENI---	97
QY	93	WNOYENLWLP-----DRMSLVSGFVEGKDEQGR	123
Db	98	WNOFN---WPFGLLCRVINGVIKANLFIISFLVVAISQDRYRVLHPMASGRQRRRQ	153
QY	124	LRRLIRYANLGNVILRSVAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHMFW	183
Db	154	ARVTCVLIVWVGGL---SIFTLRSIQVDPDLNITACI-----LLLPH----	195
QY	184	PWVFNLSMKAWLGGRIID-----PI-----LQSLNEMNTLTTCGHLYA	226
Db	196	-----EAMHPARIVELNIGLFLPLAAIVFFNVHILASLTRREVSRTCCG--R	243
QY	227	YDMISIPLVYTVVAVYSFPLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFVY-	285
Db	244	KDSKTTALITLVVA-----FLVCWAPYHF-----PAPLEFLFQVQ	279
QY	286	-----GWLKVGLSRA-----LLGWRHGQRGHGOOLLETRMOCQERK	321
Db	280	AVRGCFWEDFDLGLQLANFAFTNSLNPNVIYFVG-----RLPRTKWELYKQCTPKS	334
QY	322	VSREVS-----QAWNR	333
Db	335	LAPISSSHRRKEIFQLFWR	352

Search completed: July 10, 2003, 12:27:53
 Job time : 10.9682 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:04 ; Search time 17.1292 Seconds
(without alignments)
3283.199 Million cell updates/sec

Title: US-09-622-964-3

Perfect score: 3120

Sequence: 1 MTITVTSQVANARLGSFRL.....TLKDHDPYVALNDRDEAHS 585

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	814	26.1	612	2 T32368	hypothetical prote
2	770.5	24.7	557	2 T32367	hypothetical prote
3	758	24.3	584	2 T19565	hypothetical prote
4	752	24.1	632	2 S4917	ZK688.2 protein -
5	751	24.1	405	2 T27971	hypothetical prote
6	717.5	23.0	1355	2 T28715	hypothetical prote
7	717	23.0	499	2 T27630	hypothetical prote
8	682	21.9	413	2 T21644	hypothetical prote
9	679	21.8	450	2 T18781	hypothetical prote
10	669.5	21.5	523	2 T18782	hypothetical prote
11	656	21.0	459	2 S40708	hypothetical prote
12	646.5	20.7	513	2 T24210	hypothetical prote
13	632.5	20.3	405	2 S42371	hypothetical prote
14	622.5	20.0	613	2 T16885	hypothetical prote
15	622	19.9	420	2 B89710	protein C43G2.4 [i
16	616.5	19.8	512	2 T19806	hypothetical prote
17	585	18.8	400	2 T21670	hypothetical prote
18	583	18.7	387	2 H89192	protein F32G8.4 [i
19	495	15.9	400	2 T20050	hypothetical prote
20	470	15.1	530	2 T28037	hypothetical prote
21	455	14.6	444	2 T20048	hypothetical prote
22	447	14.3	806	2 T19468	hypothetical prote
23	441.5	14.2	411	2 T28038	hypothetical prote
24	395	12.7	434	2 T20922	hypothetical prote
25	127.5	4.1	315	2 AB0677	probable membrane
26	115.5	3.7	613	2 B82338	conserved hypothet
27	115	3.7	516	2 AE0665	probable membrane
28	113	3.6	932	2 S62555	protoplast regene
29	111.5	3.6	825	2 T47164	hypothetical prote

30	111	3.6	481	2 T42226	hypothetical prote
31	107	3.4	2606	2 T03159	large tegument pro
32	105	3.4	1707	2 T18951	hypothetical prote
33	104.5	3.3	457	2 D71505	probable transloca
34	104.5	3.3	457	2 S44484	secy protein - Chl
35	104.5	3.3	785	2 A86299	hypothetical prote
36	104.5	3.3	786	2 H64817	probable membrane
37	104	3.3	485	2 T38255	hypothetical prote
38	103.5	3.3	741	2 F90739	probable transport
39	103.5	3.3	741	2 H85589	probable transport
40	103	3.3	321	2 G90894	hypothetical prote
41	103	3.3	321	2 B85723	hypothetical prote
42	103	3.3	961	2 S58787	UFD2 protein - yea
43	102.5	3.3	321	2 C64906	probable membrane
44	102.5	3.3	1002	2 T43236	carboxypeptidase C
45	101.5	3.3	343	2 T40306	hypothetical prote

ALIGNMENTS

RESULT 1

T32368

hypothetical protein C01B12.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000

C:Accession: T32368

R:Scheet, P.; Maggi, L.

submitted to the EMBL Data Library, September 1997

A:Description: The sequence of C. elegans cosmid C01B12.

A:Reference number: Z21156

A:Accession: T32368

A:Molecule type: DNA

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-612 <SCH>

A:Cross-references: EMBL:AF025458; PIDN:AAB70976.1; GSPDB:GN00020; CESP:C01B12.3

A:Experimental source: strain Bristol N2; clone C01B12

C:Genetics:

A:Gene: CESP:C01B12.3

A:Map position: 2

A:Introns: 25/3, 60/2, 105/2, 138/3, 212/3, 319/3, 369/2, 467/2, 508/3, 573/1

C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 26.1%; Score 814; DB 2; Length 612;

Best Local Similarity 33.6%; Pred. No. 2.3e-55;

Matches 193; Conservative 98; Mismatches 221; Indels 62; Gaps 10;

QY 1 MTITVTSQVANARLGSFRLLLCWGSGIYKLYGFLIPLLCYIIRFYRLALTEEQQL 60

DB 1 MTTVTSLDVASSSFCLYKLLFRMKGSIWKSWAELVWMLCLYAVLSVIYRCLLTWKQRA 60

QY 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVLVTRWNNQYENLPWDRMLSVSGFVEGKDFQ 120

DB 61 TFEDLCIFDFYSNPFIPITFMLGFFVSAVTRWMOIFDNIGMIDTFCLMITQYIKETER 120

QY 121 SRLRLTLIRYANLGNVLILRSVSTAVYKRPSSAOLVQAGPMTAEHKLEKLSLPHNM 180

DB 121 AKCVRENCIRSYILTQANVYRDVAASVKRPTFNHLVTAGLMTKEKAEFESISPHAK 180

QY 181 FWVPWVPANISMKAWLGGRIIDPILLOSLLNEMNTLTQCCHLYAYDWISIPLYVTQV 240

DB 181 YWQPMHMLFSMITLARDEGMISDIIYVDLMKQRQRVRNLTSLTDFDVPVPLVYTVQV 240

QY 241 TVAVYSFELTCLVGRQFLNPAKAYPG---HELDLVVPVTFLOFFFYVWLNKVAQLINP 297

DB 241 HLAVRSYELIALFGROYLHPESNRINDPKQITDLYVPIMSLLQYFFITGMMKVAEVLNLP 300

QY 298 FGEDDDPFETNIVDRNLQVSLAVDENHQDLPRMPEPMYWNKPPQPPYTAASQAFRA 357

DB 301 LGEDDDDFECNWLDRNLQVSLAVDENHQDLPRMPEPMYWNKPPQPPYTAASQAFRA 360

QY 358 SFMGSTFNISLUNKMEFQPNQE-----DEDAHAGIIGFLGQ----- 397

DB 358 SFMGSTFNISLUNKMEFQPNQE-----DEDAHAGIIGFLGQ----- 397

Db 361 PQVSCADMPTEBFFMVPRRTLSRMSHWGDMEDTD---VVPVVLKHTRDNVNAS 417
Qy 398 -----SHDHPHPRANSRTKLLWPKESSLHLEGLPKNHKAQKQVRGOE 440
Db 418 GESLAPNSFANGGRKLESEMPFRWRAGSRIGDRYKRNSSAQDP--ENGMAKQNSIDENA 475
Qy 441 DNKAWKL-KAVDAPKSGPLYPQGYGAPQPLSP-----TTPMPPLEPSAPSKLHS 491
Db 476 DIHSNRLDQASGTPKSGRLWS-----SMPQTLEMLKNKFNFPVKNYNDGMDKRELQN 530
Qy 492 VTGI-DTKDKSLKTVSSGA--KKSFELLSESDGA 522
Db 531 PTPITDIDPLHVASSQSWFNESLPVKEEEA 564

RESULT 2

T32367
hypothetical protein C01B12.5 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
C/Accession: T32367
R/Scheet, P.; Maggi, L.
submitted to the EMBL Data Library, September 1997
A/Description: The sequence of C. elegans cosmid C01B12.
A/Reference number: Z21156
A/Accession: T32367
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-557 <S>
A/Cross-references: EMBL:AF025458; PIDN:AB70978.1; GSPDB:GN00020; CESP:C01B12.5
A/Experimental source: strain Bristol N2; clone C01B12
C/Genetics:
A/Gene: CESP:C01B12.5
A/Map position: 2
A/Introns: 60/2; 105/2; 212/3; 318/3; 368/2; 442/1; 509/3
C/Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 24.7%; Score 770.5; DB 2; Length 557;
Best Local Similarity 34.3%; Pred. No. 5e-52;
Matches 162; Conservative 90; Mismatches 171; Indels 49; Gaps 6;
Qy 1 MTITYTSQVANRLGSRSLLCWRGSIYKLLYGEPLIFLLCYIIRFYRLALTEQQL 60
Db 1 MTVAISLDVATSGELFTQIKVLLRWKSGVWKSISWSELIWLLCSYLSIYRIILLKAQRE 60
Qy 61 MPEKLYCDSTYQIPLIPISFVLGYFTLVVTRWQVNPDPRLMSLVSGVEGKDEQ 120
Db 61 VFEQLCTFFDTSFVPIVTFMGLGYVSVYNNRTKVDNVGMDTSALTIAQVIRTSK 120
Qy 121 SRIILRLTRYANGLNVLILRSVSTAVYKRPQSAQHLVQAGFMTPAEHKQLEKSLPHNM 180
Db 121 ARLIRENCVRYMIVQVMVFRDVSPIARRRFPPTIKHLIGAGLLTEDELTFDAITSPQSK 180
Qy 181 FWVPVW 240
Db 181 YWQPIQWLSLVTVAKDEGLIADSYLVLDIKQREFRILKILNLVDFMVPVPLVTVQV 240
Qy 241 TVAVVSFFUTCLVGRFLNPAKAYPG--HELDLVVVFVFTFLOFFYVGVWLVKVAEQLINPF 298
Db 241 NLAVRTYFVLAFGRQFLENNNNIPGAKWKIDYFIMTSLOIVFVGVWLVKVEWMLNPL 300
Qy 299 GEDDDDFETNWIIVRNQLVSLAVDEHQDLPRMEFDMWNKPEOPPYTAASQAPRRAS 358
Db 301 GEDDEDFETNWIIVRNQLVAVVQVAVVQVAVVQVAVVQVAVVQVAVVQVAVVQVAV 360
Qy 359 FMGSTFNISLNKEEMFQPNQDEEDAHAGII-----GRFLG---LQSHDHPHPRANSR 409
Db 361 MQSGSCINM-----NEADLDNGLSYVRRRSRSGFGDDDDASFSTENTKTQSN 406
Qy 410 TKLL-----W-----PKRESLHLEGLPKNHKAQKQVR 437
Db 407 ASMLPRHMQPQPNKISAAIAKFKYSEAPRRASVCVAAALNLNHRPSTDLR 458

RESULT 3

T19565
hypothetical protein C29F4.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C/Accession: T19565
R/Kershaw, J.
submitted to the EMBL Data Library, January 1996
A/Reference number: Z19143
A/Accession: T19565
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-584 <WIL>
A/Cross-references: EMBL:Z68335; PIDN:CAA92730.1; GSPDB:GN00022; CESP:C29F4.2
A/Experimental source: clone C29F4
C/Genetics:
A/Gene: CESP:C29F4.2
A/Map position: 4
A/Introns: 36/2; 78/2; 104/2; 149/2; 174/2; 221/2; 256/3; 333/3; 412/2; 484/2; 526/1; 5
C/Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 24.3%; Score 758; DB 2; Length 584;
Best Local Similarity 31.8%; Pred. No. 5e-51;
Matches 172; Conservative 110; Mismatches 195; Indels 64; Gaps 11;
Qy 19 RLLLCWRGSIYKLLYGEPLIFLLCYIIRFYRLALTEQQLMPEKLYCDSYIQLIPI 78
Db 63 KLIFRWKSGSLWKAIFYLDLVWCFYAFISVYRYALDRSQDQTFERFMQFCNRRLDYIPI 122
Qy 79 SFVLGYFTLVVTRWQVNPDPRLMSLVSGVEGDEQSRLLRRLTRYANGLNVL 138
Db 123 NFMLGFVTVINRWMTQFANLGMIDNIALFTSWYLSGNDERGRIIRSRVCMVCSQTM 182
Qy 139 ILRSVSTAVYKRPQSAQHLVQAGFMTPAEHKQLEKSLPHNMFWVWVWVWVWVWVW 198
Db 183 VFRDIHIGVRKFPPTLETWVAAGIMTSSELKKYNEVESRYAKYTWLGFNTFLLNEARRE 242
Qy 199 GRIRDPILLQSLNEMNTLRTQCHLYAYDWISIPLVYTVQVTVAVYSFELTCLVGRQFL 258
Db 243 GRIESAYTQNAEAEIRTPRSGLSLIWYDWVPIPLMYPQVPMATHCYLVCLVSRQFV 302
Qy 259 NPAKAYPGHELDLVVVPVFTFLOFFYVGVWLVKVAEQLINPFDDDDDFETNWIIVDRNLQVS 318
Db 303 INSDAVNTTIEDLGVFPFTIIEFIFVWGLWLVKVAEQLINPFDEDDDFDCNFLIDRLNLTVA 362
Qy 319 LLAVDEMHOQLPRMEFDMWNKPEOPPYTAASQAPRRASFGSTFN-----ISLNKEEME 374
Db 363 MGIVDDSHDDGPILEKDMFWN-DTVSPLYSSAAQORNVNFFYFGSATNADAQIPDDVVRQIT 421
Qy 375 FQPNQDEEDAHAGIIIGRFLGQSHDHPHPRANSRTKLLWPKESSLHLEGLPKNHKAQK 434
Db 422 MIPHPLNEK-----LDQMYGKRT--NRPP-VQSVVEL---KRQORFSTG-----459
Qy 435 NVRGQEDNKAWKLKAVDAPKSGPLYPQGYGAPQPLSPFPMFPFLEPSAPSKLHSVTG 494
Db 460 -----NNRKQLEAKFNSKLGAMFQKRKSKSLTFTPEST-----KAPASTD 501
Qy 495 IDTKDKSLKTVSSGAKKSFELLSESDGALMEHPEVQVQRKRTVEFNLTDPMPDENHLKE 554
Db 502 IET-----VSTFENTOKPCY-----SNPDCIVEVDEESQ-----DIPKVTBEPKQE 542
Qy 555 P 555
Db 543 P 543

RESULT 4

S44917
ZK688.2 protein - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C/Accession: S44917

R.Wilson, R.
 submitted to the EMBL Data Library, May 1993
 A:Description: Sequence of the C. elegans cosmid ZK688.
 A:Reference number: S44913
 A:Accession: S44917
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-632 <WIL>
 A:Cross-references: EMBL:U16621; NID:g289775; PID:g289780
 C:Genetics:
 A:Introns: 60/2; 105/2; 268/3; 366/2; 403/2; 435/2; 599/2

Query Match 24.1%; Score 752; DB 2; Length 632;
 Best Local Similarity 35.9%; Pred. No. 1.7e-50;
 Matches 152; Conservative 94; Mismatches 163; Indels 14; Gaps 4;
 QY 1 MTITTSQVANARLGSFSLLCWRGSIYKLYGEBELIFLLCYIIRFYRLALTEBOQL 60
 DB 1 MTINYNLAVSTSKPWTFLKLLKWRGSIKWKAVILELAVNLVLYGILSVIRYALNPGQOR 60
 QY 61 MFEKLTLYCDSYIOLIPISFVLGFFVYVTLVTRWNNQYENLPPDRMLSLVSGFVEGKDEQ 120
 DB 61 TPERIVQVCDRLSYIPLNFMGLGFFVAVVNRWYLYQIIGFDNIGLMAAEVVRGRTQ 120
 QY 121 SRLLRRTLIIRYANLGNVLIRSVSTAVYKRFPSAQHLVOAGFWTPAEHKQLEKLSLPHNM 180
 DB 121 AMYRNIRVRYCELAQVLFRDISMRTRRRFTLDTVVAAGFMPEKDRFDEIQYKSK 180
 QY 181 FWVPVWFWFANLSMKAWLGGRIKRDPIILLQSLNEMNTLRTOCCGHLAYADNISIPLVYTV 240
 DB 181 YWVPQWAFSLTYEARKKLIESDYVQVVDYQDEIKKFTGLAWICNYDWPIPIWYPLQVL 240
 QY 241 TVAVSVFPLTCLVGRQFLNPAKAYFPGHELDLVVPVFTFLQFFYVGLWLVKVAEQLINPFG 300
 DB 241 CLAVHTYFLVCLARQYVYVSEHADNKTIEDLYFIMSTLQIFYNGMKVAAEAMLNPFGE 300
 QY 301 DDDDFETWIVDRNLQVSLAVDEMHDQLPRMEPDMYWNKPEPQPYTAASQAFRRASF 360
 DB 301 DDDDFECNALIDRNITWLMVMDQGYDRAPLDFDWDDE-EVEPLYSEETAKIPNNPLK 359
 QY 361 GSTFNISLNK--EEMEFOPNOBDEDAHAG-----IIGRFLGLOSHDHPHPRANSR 409
 DB 360 GSVSDVKLPEYVHEIKWPHCDASPLVPGDEMRRRVSVVVPKPSDQOHHHGHG--TR 417
 QY 410 TKL 412
 DB 418 TSL 420

RESULT 5
 T27971
 hypothetical protein ZK675.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
 C:Accession: T27971
 R:Sim: M.
 submitted to the EMBL Data Library, November 1994
 A:Reference number: Z20448
 A:Accession: T27971
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-405 <WIL>
 A:Cross-references: EMBL:Z46812; PIDN:CAAB6845.1; GSPDB:GN00020; CESP:ZK675.3
 A:Experimental source: clone ZK675
 C:Genetics:
 A:Gene: CESP:ZK675.3
 A:Map position: 2
 A:Introns: 43/2; 84/2; 110/3; 174/3; 233/3; 264/2; 335/3; 372/1
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK649.5

Query Match 24.1%; Score 751; DB 2; Length 405;
 Best Local Similarity 39.2%; Pred. No. 1e-50;
 Matches 153; Conservative 79; Mismatches 140; Indels 18; Gaps 5;

QY 1 MTITTSQVANARLGSFSLLCWRGSIYKLYGEBELIFLLCYIIRFYRLALTEBOQL 60
 DB 1 MTISYSD-----TFLKLLFRWGSWKALWKHLLIFLTWYIINAYYRFGMTKEQQN 52
 QY 61 MFEKLTLYCDSYIOLIPISFVLGFFVYVTLVTRWNNQYENLPPDRMLSLVSGFVEGKDEQ 120
 DB 53 EPIKYVMLVDGWTKEIPLTFLGFFVYVAMVIRWWDCCQLISWFDHLLYVNSALIRQDPE 112
 QY 121 SRLLRRTLIIRYANLGNVLIRSVSTAVYKRFPSAQHLVOAGFWTPAEHKQLEKLSL--P 177
 DB 113 TRIIRKTIARTILTSVLARSISRLVARYPTDHLVDSGLMTKEEMVWFKXILVHVP 172
 QY 178 HNMFWVWFWFANLSMKAWLGGRIKRDPIILLQSLNEMNTLRTOCCGHLAYADNISIPLVY 237
 DB 173 HQKWVPLNWIQTMVRCFEKGTILTHNELRVLLDALEKYRNGFFQLFYDIAIPLVY 232
 QY 238 QVTVAVSVFPLTCLVGRQFLNPAKAYFPGHELDLVVPVFTFLQFFYVGLWLVKVAEQLINP 297
 DB 233 QVSTISVGYFLFALIGRQY--PSKNENEIVDVVYVPIITLQFLFYVGLWLVKVGEDLMFP 290
 QY 298 FGEDDDFETWIVDRNLQVSLAVDEMHDQLPRMEPDMYWNKPEPQPYTAASQAFRA 357
 DB 291 FGADDEDFEFYILERNLEVSMLIVDELHNQVPPYVVSLE--DDEIRLLHTSASSKLSNH 348
 QY 358 SPWGSTFNISLNKEMEFQP---NQEDBED 384
 DB 349 PQRQHLRKLKFNVDAMQVAVPGNHLNERD 378
 RESULT 6
 T28715
 hypothetical protein T21D12.9b - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T28715
 R:Woessner, J.
 submitted to the EMBL Data Library, August 1997
 A:Description: The sequence of C. elegans cosmid T21D12.
 A:Reference number: Z20514
 A:Accession: T28715
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-1355 <MOE>
 A:Cross-references: EMBL:AF016687; PIDN:AAC48095.1; GSPDB:GN00022; CESP:T21D12.9b
 A:Experimental source: strain Bristol N2; clone T21D12
 C:Genetics:
 A:Gene: CESP:T21D12.9b
 A:Map position: 4
 A:Introns: 38/2; 84/2; 132/2; 204/2; 275/2; 351/2; 519/3; 615/1; 714/1; 758/2; 786/3; 8

Query Match 23.0%; Score 717.5; DB 2; Length 1355;
 Best Local Similarity 31.6%; Pred. No. 2.4e-47;
 Matches 192; Conservative 97; Mismatches 228; Indels 91; Gaps 15;
 QY 1 MTITTSQVANARLGSFSLLCWRGSIYKLYGEBELIFLLCYIIRFYR--LALTEQ 58
 DB 787 MTVSNSDVSTSLIAFLRLQLRWGSKVWSVMELEIFSFCAITSIYTNVFLSDEQ 846
 QY 59 QLMPEKLTLYCDSYIOLIPISFVLGFFVYVTLVTRWNNQYENLPPDRMLSLVSGFVEGKD 118
 DB 847 RVFDNFAALFDQTLNYPILTFMLGFFVTIIVGRWDFINIGWVDNTALLIATVIRGSD 906
 QY 119 EQSRLRRTLIIRYANLGNVLIRSVSTAVYKRFPSAQHLVOAGFWTPAEHKQLEKLSL--P 177
 DB 907 EKSRIMRTAIRLVLTQVILFRDISLVVKRRPTLETVVASGFMLESEREKAYALSIGK 966
 QY 178 HNMFWVWFWFANLSMKAWLGGRIKRDPIILLQSLNEMNTLRTOCCGHLAYADNISIPLVY 237
 DB 967 YPKYFLPIQWCFSLLYDARAQKIGADVMLNELIKSVGDFRKLGLQLLNFDWVPIPLVYP 1026
 QY 238 QVTVAVSVFPLTCLVGRQFL-----NPAKAYPGHELDLVVP-VFTFLQFFYVYVGL 288

Db 1027 QVVFLAVRIYFFLCILIAQSVLIDGKPKDQNSVPYP-----FVPFLMTSLQFVYVYGMW 1080

QY 289 KVAEQILNPFGEEDDDDDPETNWIIVDRNLQVSLIAVDHMQDLPRMEPDYMNKPEPOPPYT 348

Db 1081 KVAESLWNLPGEDDDDDPECYNLDLDRNLAVGLAIVDS--SEAAPHYKDVGFGLAIADPLYS 1139

QY 349 AASAOQFRASFGSTGNISLNKEMEFQPNQOE--DEDAHAGIIG-----RFLGLQSH 399

Db 1140 SDTANTHINPQIGSAATYETQENIIRPHVDNDMEDGDDVEGCNPRQLSRFSVSVSV 1199

QY 400 DHPPRANRSRTKLLWPKRESLLHGLPKNHKAQKQNVGQEDNKAWKKAIVDAFKSGPLY 459

Db 1200 NRN---CESRTSL--SRNPFTIIRLSQFGSKVN-----1231

QY 460 QRGYYSAPQPLSPPTMPFPLPSPKLSHVSIGTDKSLKTSVSSGAKKSFEILLSBS 519

Db 1232 -RPG-----KLFSSQFSINTAMGDN-----DIGSCASILGELAEBS 1267

QY 520 D-----CALMEHPEVSQVRKTVFENLTDMEIPEINHLKPLEQSPNIHTLLKDHMDPTW 575

Db 1268 NKASQGLLTPEYAGSFRNTPIDMLTSVPE-----EDEEAQKTRTSVDLRKWKEMVSE 1320

QY 576 ALENRDEA 583

Db 1321 NEKKREEA 1328

RESULT 7

T27630

hypothetical protein ZC518.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000

C:Accession: T27630

R:Thomas, K.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z20396

A:Accession: T27630

A>Status: Preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-499 <WIL>

A:Cross-references: EMBL:Z68753; PIDN:CAA92989.1; GSPDB:GN00022; CESP:ZC518.1

A:Experimental source: clone ZC518

C:Genetics:

A:Gene: CESP:ZC518.1

A:Map position: 4

A:Introns: 30/3; 60/2; 106/2; 239/3; 267/3; 315/3; 329/3; 380/2; 405/3; 445/2; 466/1

C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 23.0%; Score 717; DB 2; Length 499;

Best Local Similarity 35.2%; Pred No. 6.3e-48;

Matches 160; Conservative 78; Mismatches 158; Indels 58; Gaps 9;

QY 1 MTITVTSQVANRLGSGFRLLCWRGSIYKLYGFLFLCYIIRFYIRLALTEEQOL 60

Db 1 MTISYTLDSQTNLQSFSLRLWRGSIYKLYGFLFLCYIIRFYIRLALTEEQOL 60

QY 61 MFEKLTLYCDSYIQL-IPISFVLGVFVTLVTRWQVNLPPRLMSLSGFGVEGDE 119

Db 61 VFEQIRIFDNKLDANIPLTLGLGFVGFVVARWGSILNGIWDASLLFATYIRGADE 120

QY 120 QSRLRLRTIIRYANLGNVLIILRSVAVYKRPFSQAHLVQAGFMTPAEHKOLEKLSLPHN 179

Db 121 ETRVIRNRLVYLSQALVLRDISQVQRKRPFTMDTLAASGLMTHHEMDLIDHKDYS 180

QY 180 MFWPVPVWVFWANLGMKAWGGRIRDPILLOSLLNEMNTLRTQCGHLYAYDWISIPLVYQV 239

Db 181 RYWTISQWLSNLVVEQCKGKGVDSYLLMNKIVDEIGKPRHGLASLLKYDWPVPLVYPOV 240

QY 240 VTVAVYSFELCLVGRQFL-----NPAKAYPGHELDLVVFTFLQFPFYVGLKVAEQLI 295

Db 241 IFLAVRIYFMICLIGRQFIVTGPNS-----GIDLWLPITTMVQFLVYMGWKMVAEALL 294

QY 296 NPFGEEDDDDDPETNWIIVDRNLQVSLIAVDHMQDLPRMEPDYMNKPEPOPPYT 341

Db 295 NPLGEDDDDDLECYNIIDKNLITGLSIVDTWKKHDDTGYSWVEHMAKTPAQKDEFGWID 354

QY 342 EPQPPYTAASAQFRASFGSTGNISLNKEMEFQPNQOE--DEDAHAGIIG-----RFLGLQSH 399

Db 355 KIAPLYSMESABRSVHPLVGSASKINLVKNKEIVMTPHKNKLSLDPSEQKTYLRVNV 414

QY 380 EDEDAHAGIIGRFLGLQSHDHPHPRANSRTKLL 413

Db 415 SDHNAKHA---KQKGLK-----RANSPDKCL 437

RESULT 8

T21644

hypothetical protein F32B6.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21644

R:Basham, V.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19453

A:Accession: T21644

A>Status: Preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-413 <WIL>

A:Cross-references: EMBL:Z81074; PIDN:CAB03043.1; GSPDB:GN00022; CESP:F32B6.9

A:Experimental source: clone F32B6

C:Genetics:

A:Gene: CESP:F32B6.9

A:Map position: 4

A:Introns: 22/2; 61/1; 141/2; 176/3; 200/3; 290/3; 332/1

Query Match 21.9%; Score 682; DB 2; Length 413;

Best Local Similarity 39.6%; Pred No. 2.6e-45;

Matches 141; Conservative 67; Mismatches 126; Indels 22; Gaps 7;

QY 1 MTITVTSQVANRLGSGFRLLCWRGSIYKLYGFLFLCYIIRFYIRLALTEEQOL 55

Db 1 MTISYS-----GNVIRILLRWKGSIRWTAKKELLIVLYSVRVFLKIGIDLDD 52

QY 56 EQQL-----MFEKLTLYCDSYIQLIPISFVLGVFVTLVTRWQVNLPPRLMSLS 111

Db 53 EDRLLKWRMPEFTPCRCQDSYTRILPLTLGLGFVGSVNVARWVRQFETLYWPEDLSVLC 112

QY 112 GFVEGKDEQSRLRLRTIIRYANLGNVLIILRSVAVYKRPFSQAHLVQAGFMTPAEHKOL 171

Db 113 TVLHQHDEKSKRRRTIIRYANLGNVLIILRSVAVYKRPFSQAHLVQAGFMTPAEHKOL 172

QY 172 EKLSLPH-NMFWPVPVWVFWANLGMKAWGGRIRDPILLOSLLNEMNTLRTQCGHLYAYDWI 230

Db 173 EAMAEANESSRWITPLHMIQILMRQVBEHKKPTASLNFQFVGLRIFRQSLRKLYSYDWV 232

QY 231 SIPLVYTVQVTVAVYSFELCLVGRQFLNPAKAYPGHELDLVVFTFLQFPFYVGLKVAEQ 290

Db 233 CVPLVYTVQVLAATYSFPPFTLFGQFLFP-DITGKELDLVVPVFTIVQVFLFVGFV 291

QY 291 AEQLNPFGEEDDDPETNWIIVDRNLQVSLIAVDHMQDLPRME--PDYMNKPEP 343

Db 292 QDLMRPFGDLDDDIENLYILDRNVRISPAIVNQLQSPDPFESNDDKLWHMHP 347

RESULT 9

T18781

hypothetical protein B0564.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T18781

R:Lightning, J.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z19021

A:Accession: T18781

A>Status: Preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-450 <WIL>
A:Cross-references: EMBL:Z73422; PIDN:CAA97765.1; GSPDB:GN00022; CESP:B0564.3
A:Experimental source: clone B0564
C:Genetics:
A:Gene: CESP:B0564.3
A:Map position: 4
A:Introns: 60/2; 105/2; 320/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.8%; Score 679; DB 2; Length 450;
Best Local Similarity 32.9%; Pred. No. 5e-45;
Matches 155; Conservative 80; Mismatches 184; Indels 52; Gaps 9;

QY 1 MTITYTSQVANARLGSRLLLCWRGSIYKLYGEBFLIFLLCYIIRFYRLALTEEQQL 60
DB 1 MTINYHKEIKTSHTWKFVLLPRWKGSIWKAIYMETIIPILCYIISVYVVTAMSEBSQR 120
QY 61 MPEKLTLYCDSYIQLIPISFVLGFFYTLVTRWNOYENLPWDRMLSLVSGFVEGKDEQ 120
DB 61 TFSVIRYCDKGLSPFLPFLVGGFTIIVDRWTKLWRTVGFDDVCLLANLVVGTSEK 120
QY 121 SRLRLTLIRYANLGNVLILRSYSTAVYKRFPSAQLVQAGFMTPAEHKQLEKLSLPHN- 179
DB 121 AIYRENIARYCALTOLLVFRDYSMTTRRRFPMTETVVAAGFMSKDELDLYNSYTTKNS 180
QY 180 ----MFWVPWVWFANLSMKAWLGGRIIDPILQSLNEMNTLTQCGHLYAYDWISIPLV 235
DB 181 RLKGYKIWIIPANWALCMYKARKDGYIESDYFKAQMEGEIRTWRTNIEWVCYNDVPLPLM 240
QY 236 YQVQVTVAVYSFPLCLVGRQFLNPAKAYPGHGLDLVVPVTFLOPFFYVGMKVAEQLI 295
DB 241 YPOLVCLAVNLFLVSIIRQ-LVTEKHQKQVDEVDPVFPVMTFLOPFYMGWLKVIEVML 299
QY 296 NPFGEDEDDDFETNWIIVDRNLQVLSLAVDEMHDLPMEPDYMNKPEPPOPPYTAASAOQR 355
DB 300 NPFGEDEDDDFETNALIDRNIITWGLKXVNT-MKTPELLKQDFDEVVLVSLLYSESSQIS 358
QY 356 RASFNGSTFNISLNK--EMEPQNOEDE 382
DB 359 NYHYHGSTSEVHLEQKCSVRMIPHSQE 387

RESULT 11
S40708
hypothetical protein C07A9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Aug-2000
C:Accession: S40708
R:Smith, M.
submitted to the EMBL Data Library, December 1993
A:Reference number: S40701
A:Accession: S40708
A:Molecule type: DNA
A:Residues: 1-459 <SMI>
A:Cross-references: EMBL:Z29094; NID:G436440; PID:G436448
C:Genetics:
A:Introns: 40/3; 77/2; 103/2; 109/2; 154/2; 179/2; 261/3; 411/2
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.0%; Score 656; DB 2; Length 459;
Best Local Similarity 32.3%; Pred. No. 3.2e-43;
Matches 129; Conservative 91; Mismatches 151; Indels 28; Gaps 5;

QY 1 MTITYTSQVANARLGSRLLLCWRGSIYKLYGEBFLIFLLCYIIRFYRLALTEEQ- 59
DB 44 LSYNVYDLATSKSLMIVRMIFKWRGSLWQAVYKELIVWICAYSLVSVIYRPAITRSQKE 103
QY 60 ----LMPEKLTLYCDSYIQLIPISFVLGFFYTLVTRWNOYENLPWDRMLSLVSGFV 114
DB 104 QNKBIIIIFERFEGYCDARMGYLPLNVLGFFCNIIRRLKLYTSLGNIIDNALFVSAYV 163
QY 115 EGKDEQSRLRLTLIRYANLGNVLILRSYSTAVYKRFPSAQLVQAGFMTPAEHKQLEK 174
DB 164 RGTDDRARQIRNRNIICVVISQCLVFRDIHVGVRRRPTLEAVAQAGIMLPHLEKFNIS 223
QY 175 SLPHNMFVWVWFANLSMKAWLGGRIIDPILQSLNEMNTLTQCGHLYAYDWISIP 234
DB 224 KSRYOKYVWVSNWALELLNVAKTEKSIDGNARNAIAQETSKFSALTVMYDMVPIPL 283
QY 235 YVTVQVTVAVYSFPLCLVGRQFLNPAKAYPGHGLDLVVPVTFLOPFFYVGMKVAEQ 294
DB 284 MYPOLVNMVAVHTYFPLCIFTRQPFISADAHNKTEVDLYIPPMITIERIFYMGWLKVAME 343
QY 295 INPGEDEDDDFETNWIIVDRNLQVLSLAVDEMHDLPMEPDYMNKPEPPOPPYTAAS 346
DB 344 LNFGEDEDDDFCNLLIDRLNLAIGLTSVDDAYDQLPVKPDVFTGGVSKPDLSDTSLSK 403
QY 347 -YTAASAOFRASFMGSGTFNLSLNKEMEF-----QPNQ 379

RESULT 10
T18782
hypothetical protein B0564.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T18782
R:Lightning, J.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19021
A:Accession: T18782
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-523 <WIL>
A:Cross-references: EMBL:Z73422; PIDN:CAA97766.1; GSPDB:GN00022; CESP:B0564.4
A:Experimental source: clone B0564
C:Genetics:
A:Gene: CESP:B0564.4
A:Map position: 4
A:Introns: 60/2; 105/2; 320/3; 405/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.5%; Score 669.5; DB 2; Length 523;
Best Local Similarity 36.5%; Pred. No. 3.4e-44;
Matches 142; Conservative 73; Mismatches 165; Indels 9; Gaps 4;

QY 1 MTITYTSQVANARLGSRLLLCWRGSIYKLYGEBFLIFLLCYIIRFYRLALTEEQQL 60
DB 1 MTINYHKEIKTSHTWKFVLLPRWKGSIWKAIYMETIIPILCYIISVYVVTAMSEBSQR 120


```
Db 195 VPEEISHTDDRMKYLPLTFMLGFFVTTVPERWRSALNVNPFIESVALSVAVLLPGKGRE 254
Qy 121 SRIILRETLIRYANLGNVLILRSVSTAVYKRFPSQAHLVQAGFMTPAEHKOLEKLSLPHNM 180
Db 255 DRLTRRAIRYVVLHOLIVFRDISMRVRFFPTLYKVVVDAGFWQBELDVLESVNOESSQ 314
Qy 181 -FWPFWFWANLSMKAWLGRIRIDPILLQSLNEMNLTFTQCCHLYAYDWISIPLYVTQV 239
Db 315 TTWVPINWANSALVAHQOKLIDQPTAFNNVIPAIFRVMETLIKFDIAPIPIAYPOV 374
Qy 240 VTVAVYSFELTCLVGRQFL-NPAKAYPGHELDLVVPVTFLOPFFVVGWMLKVAEQLINPF 298
Db 375 VFLAVRVYPAICLVSRQPLISDKSKT--QMDVPFIMTVLEFIFVIGWMLKVAEVLNLEL 432
Qy 299 GEDDDDFETNIVDRNLQVLSLAVDEMHDQLPRMPEMDMYN----- 339
Db 433 GEDDDDFEVNSIIDNNISRGMAIVDTTHGVHPLVDVDFSDPNLPAYSNSQIPRNLTG 492
Qy 340 -----KPEPQPPYTAASQF-----RRASFSGT----- 363
Db 493 SAAKVELAAPTDEVKIVRVNPEADPTBSRGSMFTRRNAYSLSRNKISIASNNLESPOE 552
Qy 364 --FNIS-----LNKSEMEPQNOEDEDAHAGIIGRFLGLQSHDHPHPRANSRTKLLWPK 416
Db 553 RKFNLSPMAGMLNKSTOPDRPTMETVSEHE-----PSHFYRGDRVHSDSVIPSF 603
Qy 417 RES 419
Db 604 RRS 606

RESULT 15
B88710
protein C43G2.4 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 01-Mar-2002
C;Accession: B88710
R;Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: B88710
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-420 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAB09111.1; PID:g1572760; GSPDB:GN000022; CESP:C43G2.
C;Genetics:
A;Gene: C43G2.4
A;Map position: 4
C;Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 19.9%; Score 622; DB 2; Length 420;
Best Local Similarity 31.7%; Pred.No.1.2e-40;
Matches 151; Conservative 82; Mismatches 140; Indels 104; Gaps 16;

Qy 1 MTITYTQVANARLGSFRLLCWRGSIYKLLYGELIFLLCYIIRF----- 48
Db 1 MTISYS-----GNPFLRLRWKGSIRWSWRELFLFLFYFIRFSAHPFNFTDPT 52
Qy 49 -----IYRLALTEEQLMFEKUTLYCDSVIQLIPISFVLGPFYVTLVVTWNNQYENLP 101
Db 53 DSKGYRKIFKVMCNFHE-----YTKMPLTLFLGfyvsvsvswrwrqfetr 100
Qy 102 WPDRLMSLVSGFVRGKDSRLRLRLIRYANLGNVLILRSVSTAVYKRFPSQAHLVQAG 161
Db 101 WPEDFLSILCLLLPSK--ESRPAHQIARYUNLTSALAWRDVSTKIRLPFSLRNIDAG 158
Qy 162 FMTPAEHKOLEKLSLPHNMFV--PWVWFANL-SMKAWLGGRIIDPT-LLOSLL----- 211
Db 159 LLTEKEVEKLQDINVS KAVILMLPVYTFKNLCETEPSGIRWLTLPLHWVQQLIDAEITA 218
```

```
Qy 212 -----NEMNLTFTQCCHLYAYDWISIPLYVTQVTVAVYSFELTCLVGRQPLN 259
Db 219 GRGSVNYVSVAITNELKAYRISFRLLYCHDWVCVPLVITQVAALATYSYFFCLFGRODLN 278
Qy 260 PAKAYPGHELDLVVPVTFLOPFFVVGWMLKVAEQLINPFGEDEDDDFETNIVDRNLQVSL 319
Db 279 HDDFY---SLDAFFPLFTVVQFLFFVGVFKVQDLMRPFGLDDDDDFELSVIDRNIVTSF 335
Qy 320 LAVDEMH-QDLPRMSPDMYMNKPEQPYPYTAASQFRRASFSGTFTNISLNKEEMFQPN 378
Db 336 TIVDSLQDDDDPKFBEDVFWKHNEQ-----QQHQSMFL-PRVPTSLKNGRIDLSRN 387
Qy 379 QDEDEDAHAGIIGRFLGLQSHDHPHPRANSRTKLLWPKRESLLHEGLPKNHKAOKN 435
Db 388 -----AH-----KHPPKLTYLEM-----KNQD-----PEYKRNKN 415
```

Search completed: July 10, 2003, 12:27:09
Job time : 20.4626 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:24:19 ; Search time 22.3425 Seconds
(without alignments)
3048.258 Million cell updates/sec

Title: US-09-622-964-3
Perfect score: 3120
Sequence: 1 MTITVTSQVANRLGFSRL.....TLKDHMPYALENDEAHS 585

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 segs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdb.*
2: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdb.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdb.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pdb.*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pdb.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb.*
7: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb.*
8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdb.*
9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pdb.*
10: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pdb.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pdb.*
12: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdb.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pdb.*
14: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdb.*
15: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pdb.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1397	44.8	261	9	US-09-746-783-4
2	661	21.2	314	10	US-09-768-826-54
3	539	17.3	251	10	US-09-768-826-35
4	238	7.6	99	9	US-09-764-872-429
5	111	3.6	361	9	US-10-114-893-88
6	104.5	3.3	786	10	US-09-815-242-10103
7	103.5	3.3	454	9	US-10-156-761-12378
8	99.5	3.2	516	9	US-09-866-050A-662
9	98	3.1	410	9	US-10-149-819-5
10	97.5	3.1	713	10	US-09-801-368-408
11	97.5	3.1	1384	9	US-10-226-315-2
12	97	3.1	846	9	US-10-005-691-2
13	97	3.1	1798	9	US-10-176-847-96
14	97	3.1	2061	9	US-10-036-036-181
15	96.5	3.1	465	9	US-10-023-180-44
16	94.5	3.0	1736	10	US-09-919-497-98
17	94	3.0	1247	9	US-09-908-193-2
18	93.5	3.0	624	9	US-10-156-239-24
19	93.5	3.0	624	9	US-10-199-485-24

Sequence 24, Appl
Sequence 342, App
Sequence 341, App
Sequence 2, Appl
Sequence 19, Appl
Sequence 22, Appl
Sequence 19, Appl
Sequence 12, Appl
Sequence 54, Appl
Sequence 110, App
Sequence 21, Appl
Sequence 21, Appl
Sequence 220, App
Sequence 2, Appl
Sequence 4, Appl
Sequence 21, Appl
Sequence 2, Appl
Sequence 277, App
Sequence 43, Appl
Sequence 20, Appl
Sequence 303, App
Sequence 11, Appl
Sequence 4, Appl
Sequence 3135, Ap
Sequence 8135, Ap

624 10 US-09-795-693-24
894 9 US-09-884-465A-342
900 9 US-09-884-465A-341
1885 10 US-09-920-346-2
659 9 US-10-211-962-19
980 9 US-09-908-193-22
925 3.0 US-09-908-193-19
1252 9 US-10-118-513A-12
2243 9 US-10-176-847-54
2828 9 US-10-177-293-110
2828 10 US-09-905-129-21
2828 10 US-09-991-630-21
626 10 US-09-765-272-220
2016 9 US-09-896-994-2
932 10 US-09-754-997A-4
1162 9 US-10-226-579-2
1252 9 US-09-908-193-21
1252 10 US-09-754-997A-2
898 9 US-10-043-487-277
1162 9 US-10-079-625-43
1253 9 US-09-908-193-20
1569 9 US-10-108-603-303
855 9 US-10-245-175-11
1177 9 US-10-193-692-4
2192 9 US-10-128-714-3135
2209 9 US-10-128-714-8135

ALIGNMENTS

RESULT 1

US-09-746-783-4
; Sequence 4, Application US/09746783
; Publication No. US20030044935A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallie, Edward R.
; Racie, Lisa A.
; Treacy, Maurice
; Spaulding, Vikki
; Agostino, Michael J.
; Howes, Steven H.
; Fichtel, Kim

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 231

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/746,783

FILING DATE: 21-Dec-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Milasincic, Debra J.

REGISTRATION NUMBER: 46,931

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 261 amino acids

TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-746-783-4

Query Match
Best Local Similarity 44.8%; Score 1397; DB 9; Length 261;
Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 325 MHQDLPRMPEPMYWKPEPPYTAASAPRRASFGSTFNISLNKEEMEFQPNQDEED 384
Db 1 MHQDLPRMPEPMYWKPEPPYTAASAPRRASFGSTFNISLNKEEMEFQPNQDEED 60

QY 385 AHAGIIGRFLGQSHDHPHPRANSRTKLLWPKRESLLHGLPKVHKAAQNVRGQEDNKA 444
Db 61 AHAGIIGRFLGQSHDHPHPRANSRTKLLWPKRESLLHGLPKVHKAAQNVRGQEDNKA 120

QY 445 WKLKAVDAFKSGPLYQRPYSAPOPLGPTPMFFPLEPSAPSKLHSVTGIDTKKSLKT 504
Db 121 WKLKAVDAFKSAPLYQRPYSAPOPLGPTPMFFPLEPSAPSKLHSVTGIDTKKSLKT 180

QY 505 VSSGAKSFELLESQALMEHPEVQVRKTVFNLTMPEIPENHLKEPLEQSPNTIH 564
Db 181 VSSGAKSFELLESQALMEHPEVQVRKTVFNLTMPEIPENHLKEPLEQSPNTIH 240

QY 565 TTLKDHMDPYWALENDEAHS 585
Db 241 TTLKDHMDPYWALENDEAHS 261

RESULT 2
US-09-768-826-54
; Sequence 54, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; FILE REFERENCE: P512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 54
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-54

Query Match
Best Local Similarity 21.2%; Score 661; DB 10; Length 314;
Matches 136; Conservative 31; Mismatches 53; Indels 32; Gaps 6;

QY 172 EKLSPHNMFVPTWFEANLSMKWLGRIIRDPILLOSLLNEMWTLRTQCGHLXAYDWTIS 231
Db 13 ESKSDFNKYVPCVWFNTLAAQARDGRIRDIALLLESLNKYRAKCSMLFHYDWTIS 72

QY 232 IPLVYTVQVTVAVYSFFLTCLVGRFLNP-----AKAY-----PGHE-----LDLWVPVF 276
Db 73 IPLVYTVQVTVAVYSFFLTCLVGRFLNP-----AKAY-----PGHE-----LDLWVPVF 132

QY 277 TFLQFFVTVGKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLAVDEMHDQDPRMSPDM 336
Db 133 TLLQFFVTVGKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLAVDEMHDQDPRMSPDM 192

QY 337 YWKKPEPPPYT-AASAPRRASFGSTFNISLNKEEMEFQPNQDEEDAHAG----- 388
Db 193 YWDEDPQPPYTVATAESLRPSFLGSTFNLRMSDD-----PEQSLQVEASPGSRPAPA 247

QY 389 -----IIGREFGL 396
Db 248 AQTPLGLGRFLGV 259

RESULT 3
US-09-768-826-35
; Sequence 35, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; FILE REFERENCE: P512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 35
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-35

Query Match
Best Local Similarity 17.3%; Score 539; DB 10; Length 251;
Matches 111; Conservative 26; Mismatches 31; Indels 32; Gaps 6;

QY 224 LYAYDWISIPLVYTVQVTVAVYSFFLTCLVGRFLNP-----AKAY-----PGHE----- 268
Db 2 LFHYDWISIPLVYTVQVTVAVYSFFLTCLVGRFLNP-----AKAY-----PGHE----- 61

QY 269 LDLVVPVFTFLQFFYVGVKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLAVDEMHDQ 328
Db 62 PMYVPLTLLQFFYVGVKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLAVDEMHDQ 121

QY 329 LPRMEPDMYWKPEPPYTT-AASAPRRASFGSTFNISLNKEEMEFQPNQDEEDAHAG 387
Db 122 LPPAEKQYWDQDQPPYTVATAESLRPSFLGSTFNLRMSDD-----PEQSLQVEASPG 176

QY 388 G-----IIGREFGL 396
Db 177 GSGRPAPAAQTPLGLGRFLGV 196

RESULT 4
US-09-764-872-429
; Sequence 429, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 429
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-872-429

Query Match
Best Local Similarity 7.6%; Score 238; DB 9; Length 99;
Matches 46; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFSLLCWRGSIYKLLYGFLIFLCYIYFIRYLRALTEQQL 60

Db 13 MTVSYTKVAERFGGSGLLLEWRGSIYKLLYKEFLFGALYAVLSITVRLLLTQORY 72
 Qy 61 MFPEKLTCDSYQLPISPVLG 83
 Db 73 VIAQVARYCNRSADLPLSPVLG 95

RESULT 5

US-10-114-893-88
 ; Sequence 88, Application US/10114893
 ; Publication No. US20020193567A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: LaVallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Bowman, Michael R.
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Carlin-Duckett, McKeough
 ; APPLICANT: Kelleher, Kerry S.
 ; APPLICANT: Genetics Institute, Inc.
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
 ; FILE REFERENCE: GI 6000-10A
 ; CURRENT APPLICATION NUMBER: US/10/114,893
 ; CURRENT FILING DATE: 2002-04-02
 ; EARLIER APPLICATION NUMBER: 09/413,232
 ; EARLIER FILING DATE: 1999-10-06
 ; NUMBER OF SEQ ID NOS: 321
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 88
 ; LENGTH: 361
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (273)
 US-10-114-893-88

Query Match 3.6%; Score 111; DB 9; Length 361;
 Best Local Similarity 24.4%; Pred. No. 0.15;
 Matches 64; Conservative 35; Mismatches 97; Indels 66; Gaps 10;
 Qy 341 PEOPPYTAA-----SACFRASFMGSTFNI-----SLNKEEMEFQPNQDEDEDAHAGIG 391
 Db 107 PETQPLMLSTADKSSDSSPERASASQSTKLRPSLQKPSI----- 148
 Qy 392 RFLGLQSHDHPHPRANSTKLLMPKR--BSLLHEGLPKNHKAQKQNVRGQED-----NKA 444
 Db 149 -----FNSAGKLTSPSPKRSKSKSKSLKSHLEKAHKGSLEKCLKDYA 197
 Qy 445 WKLKAVDAFKSGPLQYRPGYSAQTPLSPTMPFPPLPSAPSLSVGTGIDTKDKSLKT 504
 Db 198 CKLASSDKPVAPPQFLKPLYSHPQNEISPSKPGPQSLAKPKPHFN-----PKRS 248
 Qy 505 VSSGAKSKSFLLSSSDGALMEHPVSQVRRKTVENPLTDMPEIPENHLKEPLEQSPPTNIH 564
 Db 249 VSLGRA--rALLSNSLA-----ETCQPKKKXKLVAKYRPLV--NDISEAKEKNTQNLH 298
 Qy 565 TT--LKHMDPYWALENRDEAH 584
 Db 299 VSSKVKSSRSFRKLDGRNNAY 320

RESULT 6

US-09-815-242-10103
 ; Sequence 10103, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-03-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10103
 ; LENGTH: 786
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-815-242-10103

Query Match 3.3%; Score 104.5; DB 10; Length 786;
 Best Local Similarity 21.1%; Pred. No. 1.9;
 Matches 69; Conservative 51; Mismatches 134; Indels 73; Gaps 16;
 Qy 31 LLYGBFLFLCYIIRIYRL-----ALTEBQQLMFEKLT-----YC 69
 Db 231 MIIGAFTIDLLALLTLFVGQVLSNLAGSRTTAFQOSLFLNAPALIEFFKAVRLRIFC 290
 Qy 70 DSYIQLIPISFVLGPFYVTVVTRMNQYENLPWDRMLSLVS-GFVEGKDEQSRLRLRT 128
 Db 291 PNVAELRE-----FTIQDESARYWSR--RLSW---LSSLYGLIVAVPIISQVN--- 336
 Qy 129 IRYANLGNVLLRSVST-AVYKRPFSQAHLVQAGFMTPAEHKQLEKLSLPHNMFVWPVW 187
 Db 337 VQIGALANVIIMLCWTVWALYIFRNKKEITQ-HLLNPAEH-SLAFFSLFIRAFALVWHM 394
 Qy 188 PANL-----SMKANLGGRIIRDPILL-----QSLNEMNTLRTQCG 222
 Db 395 LASAYFIVLPFSLFDPGNSLKFMMGATVRSIAIGIAAFVSGMFSRWLAKTITLSPTHQ 454
 Qy 223 HLY-----AYDMWISIPLVTVVTVVAVYSFPLCLVGR-QFLNPAKAYPGHE-LDLVVP 274
 Db 455 RNYPELQKRLNGWLSAALKATARIITVCVAVMILLISANGLFDVFWNLQNGAGQKTVDIR 514
 Qy 275 VFTPLQPFVYVWMLKVAEQLNPFEGD 301
 Db 515 IALIL-EFSAVGWTVLASLIENRLASD 540

RESULT 7

US-10-156-761-12378
 ; Sequence 12378, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI

```

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12378
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12378

Query Match      3.1%; Score 103.5; DB 9; Length 454;
Best Local Similarity 23.4%; Pred. No. 1;
Matches 72; Conservative 38; Mismatches 101; Indels 97; Gaps 18;

Qy 9 VANARIGSFRLLCW-RGSIYKLLYGEFLIFLLCYIIRFYRLAUTEQQLMFEKLT 67
Db 139 LAVASVGIYIVLAGSSGSTYPLGG-----LRSCQMIS--YEIAM-----GRAFASVFL 188
Qy 68 YCDS-----YIQLPISFVLGYVTLVVRWNNQYENLPM--PDRMLSLVS 111
Db 189 YSGSMSTSTVEQOQDRWYIVLLPVSVFI--YIVTMV-----GETNRAPFDPMPSEGLVG 242
Qy 112 GFVEGKDEQSRLRLRLIRYANLGNVILRSVSTAVYKRPSPSAQLVQAGFMTPAEHKQL 171
Db 243 GF-----NTEYSISKFMFMFLAEVYVNMVTSVAVSTTLF-----LGMWRAP----- 282
Qy 172 EKLSLPHNNFW-----PWVFA-----NLSMKAWLGG-----RIRDPILLO----- 208
Db 283 -----WPISTFEGANHGWMFLVFKVQVLLFFFIWLRGTLPRVYDQMLKLGKVLIP 338
Qy 209 -----SLNEMNTLTQCHLYADWISIPLYVTVVTVVAVYSPFLTCLVGRQFINPAK 262
Db 339 VSVVWMLVATVTKLRNE-----NYDFADIAL-----YVGGVLLVLLLSFVADMPREKSK 389
Qy 263 --AYPGHE 268
Db 390 EAAAPAE 397

RESULT 8
US-09-866-050A-662
; Sequence 662, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 662
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-662

Query Match      3.2%; Score 99.5; DB 9; Length 516;
Best Local Similarity 22.0%; Pred. No. 2.9;
Matches 54; Conservative 30; Mismatches 105; Indels 57; Gaps 8;

; APPLICANT: PRMEPDYMNKPEQPPYTAASQAQFRASFMGSTFNISLNKEMEFQPNQDEDEDAHAGI 389
; APPLICANT: PALQSTSKQPTPKFTALVTRGRTHKPTSEGLE---SVGPVAPDFEPTSTDLATSKV 157
; APPLICANT: IGRFLGLQSH--DHHPPRANSRTKLLWPKRESILHGLPKNHKAAKQNVRGQEDNKAWKL 447
; APPLICANT: TGOSLTQSSPVASASVSTTPKPPVPIAQLTLEPVQTSRORRRATKGGSR--- 214
; APPLICANT: KAVDAFKSGPLYQRPQYYSAPQTPLSPTPMFFLEPSAPSKLHSVTGIDTKOKSLKTVSS 507
; APPLICANT: -----APVGPKSYSYPAEB-----EPQSSASQSS 238
; APPLICANT: GAKKSPFLLSESGALMEHPEVSQVRKTVENLTDMPENHNLKEPLEQSTNHTTL 567
; APPLICANT: 239 GA-----SEADSPHQKRP-ROVTKTIVVVKGEDFGEI---QVKE---EPQETAISTPG 285
; APPLICANT: 568 KDHMDP 573
; APPLICANT: 286 KRKRDP 291

RESULT 9
US-10-149-819-5
; Sequence 5, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dying Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 1689337CD1
US-10-149-819-5

Query Match      3.1%; Score 98; DB 9; Length 410;
Best Local Similarity 23.5%; Pred. No. 2.7;
Matches 72; Conservative 25; Mismatches 133; Indels 76; Gaps 15;

Qy 330 PRMEPDYMNKPE-----PQPPYTAASQAQFRASFMGSTFNISLNK 370
Db 98 PEKEERRVWTWPMNAVALKPVLOOSREARDELPGAPPVLCSSSDLSLLGPF-----Q 152
Qy 371 EEMEFQPNQDEDEDAHAGIIGRF--LGLQSHDHHPPRANSRTKLLWPKRESLLHGLPKN 428
Db 153 SQHSFQP-LEPKPDLTSSTAGAFSALCAFPDHAERPFPEED--PGPDGEGLLKQGLPPA 210
Qy 429 HKAQKQNVRCQ-----EDNKAWKLKAVDAF--KSGPLYQRPQYYSAPQTPLSPTMFFP 480
Db 211 QLEGLKNFLHQLLETVPQNNEN---PSVDLLPKPSGFL--TVPSWEAPQVPRIPPPVKHT 266
Qy 481 LEPSA-----PSKLHSTVTGIDTKOKSLKTVSSGAKSKFELLSESD----- 520
```


Db 267 KVLAMASSLFRVPEPPSSHS-QSGSPSGSPERGGLTFRQLMEVSQLRLYQARGW 325
QY 521 GAL-----MEHEVSVQRKTVFNLTD--MPEIPENHLKEPLEQSPNTIHTTL 567
Db 326 GALPAEDLLLYLKLRLHSGRTDGRDNVPRNTDSRLGEI--PSQAVPRRLATAP 383
QY 568 KDHMDP 573
Db 384 KTEKPP 389

RESULT 10

US-09-801-368-408
; Sequence 408, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 408
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-408

Query Match 3.1%; Score 97.5; DB 10; Length 713;
Best Local Similarity 23.1%; Pred. No. 7;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
QY 349 AASAFRRASFMGSTFNISLNKEEMFQPNQDEE-----DAHAGI 389
Db 18 AIRQEFQVQSBANTYRLQ-NOKDYDFKNNQQLAEMQIIRNTVYELETHRRKMKDAYEE 76
QY 390 IGRF-LGLQSHDHPHPRANSRTKLLWPRESLLHGLPKNHKAAKONVREGQEDNKAWLK 448
Db 77 IGHKLGLQSHD-----QIASLTQV---QORQQQQQQVQOHHQQQQQQLA 120
QY 449 AVDAKSGPLYQR-PGYVSAPQTPISPT-----PMFFPLEPSAPSKLHSGVTGIDTKDKSLK 503
Db 121 AASA--SVEVAQPPATTSATPAANTTTGSPSAFFVQASRP-----NLVGSQLPITLIP 174
QY 504 TVSSGAKSKFELLSSDGNALMEH---PEVSQVRKTVFN-----LTDMPFI--PENH 551
Db 175 VVSSNAQQQLPQQQLQQOQLOQQPPQVSVAPLSNTAINGSPTSKEITTLPSVKAPEST 234
QY 552 LKEPLEQSPNTIHTT 566
Db 235 LK---ETEPENNTS 246

RESULT 11

US-10-226-315-2
; Sequence 2, Application US/10226315

Publication No. US20030120051A1
; GENERAL INFORMATION:
; APPLICANT: Poles, Elijor
; TITLE OF INVENTION: CASPR/190, A FUNCTIONAL LIGAND FOR RPTP-BETA AND THE
; TITLE OF INVENTION: AXONAL CELL RECOGNITION MOLECULE CONTACTIN
; FILE REFERENCE: 7683-111
; CURRENT APPLICATION NUMBER: US/10/226,315
; CURRENT FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: 08/826,134
; PRIOR FILING DATE: 1997-03-26
; PRIOR APPLICATION NUMBER: 60/014,199
; PRIOR FILING DATE: 1996-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-226-315-2

Query Match 3.1%; Score 97.5; DB 9; Length 1384;
Best Local Similarity 20.5%; Pred. No. 19;
Matches 95; Conservative 50; Mismatches 180; Indels 139; Gaps 21;
QY 203 DPILQSLNEMNTLTQCGHLYAYDWISPLVYTVQV-----TVAVYSFF---LTCV 253
Db 725 DPALYCNCDADQPOWRTDKGLTFFVDHLPV---TQVIGDTNRTSEAQFFLRPLRCYG 780
QY 254 GROPLNPAKAYG-----HELDLVVPVFTFLQFFP-----YVG 286
Db 781 DRNSWNTISFHTGAALRFPPIRANHSLD-----VSFVFTSAPSGVLENMGGFYCQ 832
QY 287 W-----LKVAEQLINPF---GEDD-----DDEFTN---WIVDR---NLQVSLAVD 323
Db 833 WRPPYRVVELNTRDVFADVGVNGDENLTVHSDDEFNDEHVLVRAEINVKQARLVD 892
QY 324 EMHQDLPRMEPDYMNKPEFPPTAASQAQFRRASFMGSTFNISLNKEEMFQPNQDEE 383
Db 893 HRPVLRPMPLOTYIWMEDQPLY-VGSALKRRPVGCLRAMRLNGVTL-----NLEGRA 947
QY 384 DAHAGIIGRFLGLQSHDH-----HPPRANSRTKLLWPRESLLHGLPKNHKAAKONVREGQ 439
Db 948 NASEGTSNCTGHCAPRLPCFPHGRCVERYSYTYTCDCLTAFDGPYCNH-----DIGGF 1002
QY 440 HONKAWK-----LKAVDAFKSG--PLYQRPVY-----YSAPQT 470
Db 1003 PFGTWMRYNLOSALRSAREFSHMLSRPVPQYEPGYIPGYDTPGVYHGGYRLPDY 1062
QY 471 PLSPTEPMFFPLEPSAPSKLHSGVTG-----IDTKDKSLKTVSSGAKSKFELLSSDGA 522
Db 1063 PRPGRPV-----PGYRGVYNTVGEVSFSFSTSSAPAVLLYVSVFVRDYMVAVLKDDGT 1117
QY 523 LMEHEVSVQRKTVFNLTDMPFIENHLKEPLEQSPNTIHTT 566
Db 1118 LQ-----LRYQLGTSPPYVYQLTTRPVTGQPHSINIT 1149

RESULT 12

US-10-005-691-2
; Sequence 2, Application US/10005691
; Publication No. US20020183250A1
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: MICHALOVICH, DAVID
; TITLE OF INVENTION: NOVEL USE
; FILE REFERENCE: GH-30003-D1
; CURRENT APPLICATION NUMBER: US/10/005,691
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 09/107,847
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: EP 97304996.8
; PRIOR FILING DATE: 1997-07-08
; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 846
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-005-691-2

Query Match          3.1%; Score 97; DB 9; Length 846;
Best Local Similarity 25.6%; Pred. No. 10;
Matches 45; Conservative 28; Mismatches 57; Indels 46; Gaps 10;

QY 416 KRESLLHEGLPKNHKAQNVGOEDNKAWL---KAVDAFKSGPL-----YQPCYYS 466
Db 386 RRELGHIESLPET--AADKQDSGSDNRINTVLSKEALERFDHSPTPSASRSRKSST 443
QY 467 APQTPLSPTMFFPLEPSAPSKLHSTGIDTKOKSLKTVSSGAKKSPILLSSESDGALMEH 536
Db 444 AVSDP-SSTPTKPTDSTTPROH---LPAHEKQVQRRSPSPSQSIN--SQSVGSLSLQ 496
QY 527 PEVSQVRKTVFENLTDMPRIPE-----NHLKEPLEQS-----PTNIH 564
Db 497 PVMSQA-----TNLP-IPQMSQFSAQLGAMQHLKDQLEQRTMRMEANIH 542

RESULT 13
US-10-176-847-96
; Sequence 96, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 1798
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-96

Query Match          3.1%; Score 97; DB 9; Length 1798;
Best Local Similarity 19.5%; Pred. No. 32;
Matches 105; Conservative 73; Mismatches 188; Indels 172; Gaps 29;

QY 102 WPDRLMSLVSGFVEGKDEQSRLRLRLRYANLGNVLILRSVSTAVYKRPSPSAQHLVQAG 161
Db 494 WLDKLMQLTE---EPQNSMPDII-----IWMIRGEKRLAYARIPAHQVL--- 534
QY 162 FMTPAEHKQLEKLSLPHNMF-----VPMWVFANLSMKAWLGGRIRDPILLQSL 210
Db 535 YSTSGENASGKVGKGTQTFILKYPOEKNGPKVP-----VELRVNIWLG-----LSAV 582
QY 211 LNMENLRTQCHLYAYDWISIPLVTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELD 270
Db 583 EKKFNSFAEGTFTVFAEMYNQALMFGKWT-----SGLVGR-----HKFS 623
QY 271 LVVPVFTFLQFFFY--VGLWKVAEQLINPFGEDDDDDFETNWIVD--RNLQVSLLA----- 321
Db 624 DVTGKIKLKRFFLPKGM-----EWEGWIVDPERSLLTADAGHTEF 667
QY 322 VDEMHDQLPRMPEPMYNNKPEPPYTAASAFRRASFMGSTFNISLNKEEMEFOPNOED 381
Db 668 TDEVYQNESRY-PGGDW-KP-AEDTYTDANG-----DKAASPELTCPPGHEW 712
QY 382 BEDAHAGIIGRL-----GLQSHDHPHPRANSRTKLLWPKRESLLH-----EGLPKNHK 430
Db 713 EDDAWSYDINRAVDKGEWGYGITIPPDHKPKS-----WVAAEKMYHTRRRRLVRKRK 765
QY 431 -----AAQNVRGOE---DNKAWLKAVDAFK-----SGPLYQRPGYYSAPQTPLSPTP 476
```

```
Db 766 KDLTQTASSTARAMELQOEGWEYASLIGWKHWKQSRSDTFRRRW----- 813
QY 477 MFFPLEPSAPSKLHSTGIDTKOKSL--KTVSSGAKKSPILLSSESDGALM--EHPEVSQV 532
Db 814 ----RRKMAPSETHGAAAIKLEGALGADTTEDGDEKSLKQKHSATTVFGANTPIVSCN 869
QY 533 RRKTVFNL-----TDMPEIPENHLKEP-----LEQSPTN--IHTTLKDHMDPYW 575
Db 870 FDRVYIHLRCVYQARNLLALDKDSFDPYAHICFLHRSKTTETIHTSL-----NPTM 923

RESULT 14
US-10-060-036-181
; Sequence 181, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiaug, Yuguu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 2061
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-181

Query Match          3.1%; Score 97; DB 9; Length 2061;
Best Local Similarity 19.5%; Pred. No. 39;
Matches 105; Conservative 73; Mismatches 188; Indels 172; Gaps 29;

QY 102 WPDRLMSLVSGFVEGKDEQSRLRLRLRYANLGNVLILRSVSTAVYKRPSPSAQHLVQAG 161
Db 757 WLDKLMQLTE---EPQNSMPDII-----IWMIRGEKRLAYARIPAHQVL--- 797
QY 162 FMTPAEHKQLEKLSLPHNMF-----VPMWVFANLSMKAWLGGRIRDPILLQSL 210
Db 798 YSTSGENASGKVGKGTQTFILKYPOEKNGPKVP-----VELRVNIWLG-----LSAV 845
QY 211 LNMENLRTQCHLYAYDWISIPLVTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELD 270
Db 846 EKKFNSFAEGTFTVFAEMYNQALMFGKWT-----SGLVGR-----HKFS 886
QY 271 LVVPVFTFLQFFFY--VGLWKVAEQLINPFGEDDDDDFETNWIVD--RNLQVSLLA----- 321
Db 887 DVTGKIKLKRFFLPKGM-----EWEGWIVDPERSLLTADAGHTEF 930
QY 322 VDEMHDQLPRMPEPMYNNKPEPPYTAASAFRRASFMGSTFNISLNKEEMEFOPNOED 381
Db 931 TDEVYQNESRY-PGGDW-KP-AEDTYTDANG-----DKAASPELTCPPGHEW 975
QY 382 BEDAHAGIIGRL-----GLQSHDHPHPRANSRTKLLWPKRESLLH-----EGLPKNHK 430
Db 976 EDDAWSYDINRAVDKGEWGYGITIPPDHKPKS-----WVAAEKMYHTRRRRLVRKRK 1028
QY 431 -----AAQNVRGOE---DNKAWLKAVDAFK-----SGPLYQRPGYYSAPQTPLSPTP 476
Db 1029 KDLTQTASSTARAMELQOEGWEYASLIGWKHWKQSRSDTFRRRW----- 1076
QY 477 MFFPLEPSAPSKLHSTGIDTKOKSL--KTVSSGAKKSPILLSSESDGALM--EHPEVSQV 532
Db 1077 ----RRKMAPSETHGAAAIKLEGALGADTTEDGDEKSLKQKHSATTVFGANTPIVSCN 1132
QY 533 RRKTVFNL-----TDMPEIPENHLKEP-----LEQSPTN--IHTTLKDHMDPYW 575
```

```

RESULT 15
US-10-029-180-44
; Sequence 44, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeffry C.
; APPLICANT: Trueheart, Josh
; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029,180
; CURRENT FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257,431
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-44

```

Query Match	3.1%; Score 96.5; DB 9; Length 465;
Best Local Similarity	21.2%; Pred. No. 4.5;
Matches	91; Conservative 56; Mismatches 144; Indels 139; Gaps 25;
QY	154 AQLVQAQFMTPAEHKOLEKLSLPHNMFVWVWFANLSKMWLGG--RIRQPILOS 210
DB	89 ASHAAQMGFARGAQMQQ-QQLHQNH-----GRLAETKAGGVKTRIRD-VWKHNL 137
QY	211 LNEWTLRTQCGHLIYADWHSIPLVYQVTVVAVYSF-----FLTCLVG----- 254
DB	138 AQEMAILRQ-----VEKYPYISMDTEFGIVARPIGAFTNKADYHQLRCNVDLLKMIQL 194
QY	255 --ROFLNPAKAYPGCHELDL-----VVPVFTFLOQFFVYGVMLKVAEQLINPGEDDDD 304
DB	195 GITLFSSEGEVPPNATDANGQPLGNGLVPAPCTWQFN-----RPSUEDDM 241
QY	305 FETNWTVDRLQVSLLAVDHMQDLPRMEP-----DMYWNKPEPOPPYT 348
DB	242 YAO-----ESTAMLAKAGIDPFAMHDKNGIDPFEGALLISSGLVLDDVHW----- 287
QY	349 AASAQFRRAFSMGSTFNISLNKEEMEFQPNQDEEDAHAGIIGRFGL-----QSHD----- 400
DB	288 ---VSPHSGVDFGLMKIMLCK-----PLPENESEFH-----RLNIFPFSYDIKYL 333
QY	401 HPPPRANSTRKL-----LWPKRESLLHGLPKNHAKAQNVRGQSDNKAWLKAV----- 450
DB	334 KHAGRNQAASGLQDIADGLGVKRVGIAHQA-----GSDSLVTGE---IYWKMRQLVFNG 384
QY	451 ----DAFKSGPLY-----ORQG--YYSAP-CTP-LSPTPMFPPL-EPGAPSKLHSVTGIDTK 498
DB	385 NIDEAKYSGQIWLGNQMPALLYSMQPHOTPNLNGATIYSATGTPTPNAVHSVTGSHTP 444
QY	499 DKSLKTVSSG 508
DB	445 QHALTPGATG 454

Search completed: July 10, 2003, 12:28:57
Job time : 24.3425 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	100	3.2	2183	3	US-08-746-111-5	Sequence 5, Appli
2	99.5	3.2	713	1	US-08-190-802A-63	Sequence 63, Appl
3	99.5	3.2	713	4	US-08-477-346-63	Sequence 63, Appl
4	99.5	3.2	713	4	US-08-473-089-63	Sequence 63, Appl
5	99.5	3.2	713	4	US-08-487-072A-63	Sequence 63, Appl
6	99.5	3.2	713	4	US-09-108-857-3	Sequence 3, Appli
7	99	3.2	342	3	US-08-785-928-1	Sequence 1, Appli
8	99	3.2	342	3	US-08-728-603-17	Sequence 17, Appl
9	98.5	3.2	681	2	US-08-372-255-6	Sequence 6, Appli
10	98.5	3.2	681	2	US-08-272-255-9	Sequence 9, Appli
11	98.5	3.2	681	5	PCT-US95-08565-6	Sequence 6, Appli
12	98.5	3.2	681	5	PCT-US95-08565-9	Sequence 9, Appli
13	98	3.1	192	4	US-09-134-001C-4329	Sequence 9, Appli
14	97.5	3.1	1384	4	US-08-826-134-2	Sequence 4329, Ap
15	97	3.1	846	3	US-08-885-291-55	Sequence 2, Appli
16	97	3.1	846	3	US-09-107-847-2	Sequence 55, Appl
17	97	3.1	846	4	US-09-496-672-55	Sequence 2, Appli
18	95.5	3.1	681	3	US-08-964-268-5	Sequence 55, Appl
19	95.5	3.1	870	4	US-09-134-001C-4959	Sequence 5, Appli
20	95	3.0	1139	4	US-09-046-158A-22	Sequence 4959, Ap
21	94.5	3.0	655	2	US-08-372-255-18	Sequence 22, Appl
22	94.5	3.0	655	5	PCT-US95-08565-18	Sequence 18, Appl
23	93.5	3.0	300	4	US-09-395-689-1	Sequence 18, Appl
24	93.5	3.0	765	2	US-08-663-112-2	Sequence 1, Appli
25	92.5	3.0	659	4	US-09-562-737-19	Sequence 2, Appli
26	92	2.9	626	4	US-08-961-083-220	Sequence 19, Appl
27	92	2.9	1132	2	US-08-567-508C-2	Sequence 220, App
28	92	2.9	1132	2	US-08-567-508C-2	Sequence 22, Appli

Db 908 PSELIPKQKTSKFLNRWRVASEKSGVEIIAANGEDTDVXLTNSPQONITV----- 962
 Qy 323 DEMHQDLRMEPMYWN---KPEPPOPYTAASQAFRRASFMGSGTFNISLNKEEMEPQPNQ 379
 Db 963 -----PRGESTHTNTRKPSDLPTFGVGHK-----SPHVRQEEENSFGQKRQ 1006
 Qy 380 -----EDEDHAGIIGR-FLGLOSHDHP---PRANSRTKLLWPKESSLHEGL 425
 Db 1007 LFIRTRKKKXKXKALHSPSRGFDPLRGHNHSPDRRLNHSLLHKSNETALSDDL 1066
 Qy 426 PKNHKAQKQVRGQDNKAWLKAVDAFKSGPLYQRPGYYSAPQTPLSPTMFFPLEPSA 485
 Db 1067 NOTSPMSMSTRSLPDYQYKNDTEQMSSSLDIQ-----SVPAEHSPT---FPAQD--- 1116
 Qy 486 PSKLHSTV-----GID-----TKOKSLKT-----VS 506
 Db 1117 PDQTHSTDPYSRSPPELSQLDYLSDHDFYDDIGLTSFFPDQSQKSFSSDDDDQAIP 1176
 Qy 507 SGAKSFELLSSDGMALMEHPEVSQV-----RRKTVEENLTDMPETPENHLK 553
 Db 1177 SSDLSLFTISPDLDTII-YPDLQLLSPEDNQKTSPPDLGGQVPLSPDQK 1228

RESULT 2

US-08-190-802A-63
 ; Sequence 63, Application US/08190802A
 ; Patent No. 5519003
 ; GENERAL INFORMATION:
 ; APPLICANT: Mochly-Rosen, Daria
 ; APPLICANT: Ron, Dorit
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
 ; TITLE OF INVENTION: Thereof
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dehlinger & Associates
 ; STREET: P.O. Box 60850
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-0850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 01-FEB-1994
 ; APPLICATION NUMBER: US/08/190,802A
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fabian, Gary R.
 ; REGISTRATION NUMBER: 33,875
 ; REFERENCE/DOCKET NUMBER: 8600-0139
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 324-0880
 ; TELEFAX: (415) 324-0960
 ; INFORMATION FOR SEQ ID NO: 63:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 713 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: TUP1, Fig. 46
 US-08-190-802A-63

Query Match 3.2%; Score 99.5; DB 1; Length 713;
 Best Local Similarity 23.1%; Pred. No. 0.43;
 Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;

Qy 349 AASQFRASFMGSGTFNISLNKEEMEPQPNQDEE-----DAHAGI 389
 Db 18 AIRQFLQVQSBQATYRLQ-NQKDYDFKXNQQLAEMQIQINTVYELTHRKMKADEAE 76
 Qy 390 IGRF-IGLOSHDHPHPRANSRTKLLWPKESSLHEGLPKNHKAQKQVRGQDNKAWLK 448
 Db 77 IKHLKLGLEQRDH-----QIASLTQV-QQQQQQQQQVQOHLQQQQQOLA 120
 Qy 449 AYDAFKSGPLYQRPGYYSAPOTPLSPT-----PMPEPLPSAPSKLHSTVGTIDTKOKSLK 503
 Db 121 AASA--SVPAQPPATTSATATPAANTTTGSPSAFPVQASRP-----NLVGSOLPTTLP 174
 Qy 504 TVSSGAKSKFELLSSDGMALMEH---PEYQVRRKKTVEFN-----LTDMPET---PENH 551
 Db 175 VVSSNAQQQLPQOQLQQOQLQQOQLQQOQLPPQVSVAPLSNTAINGSPTSKETTLPLSVKAPEST 234
 Qy 552 LKEPLEQSPNTIHHT 566
 Db 235 LK---ETEPENNTS 246
 RESULT 3
 US-08-477-346-63
 ; Sequence 63, Application US/08477346
 ; Patent No. 6262023
 ; GENERAL INFORMATION:
 ; APPLICANT: Mochly-Rosen, Daria
 ; APPLICANT: Ron, Dorit
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
 ; TITLE OF INVENTION: Thereof
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morrison & Foerster
 ; STREET: 2000 Pennsylvania Avenue, NW
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20006-1812
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,346
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/487,072
 ; FILING DATE: 07-JUN-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: MURASHIGE, KATE H.
 ; REGISTRATION NUMBER: 29,959
 ; REFERENCE/DOCKET NUMBER: 2550-0025.20
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 887-1500
 ; TELEFAX: (202) 887-0763
 ; INFORMATION FOR SEQ ID NO: 63:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 713 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: TUP1, Fig. 46
 US-08-477-346-63

Query Match 3.2%; Score 99.5; DB 4; Length 713;
 Best Local Similarity 23.1%; Pred. No. 0.43;
 Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;

QY 349 AASAFRRASFMGSTFNISLNKEEMFQPNQDEE-----DAHAGI 389
DB 18 AIRQEFLOVSQEAITYRLQ-NOKDYDFKMNQQLAEMQOIRNTVYELELTHRKMKAAYEAE 76
QY 390 IGRF-IGLQSHDHPHPRANSRTKLLWPKRSLHLEGLPKNHKAQKONVRQEDNKAWK 448
DB 77 IKHLKLGLEQRDH-----QIASLTQV--QOQOQOQOQOQOHLQOQOQOLA 120
QY 449 AVDAFKSGPLYQOR-PCYYSAPOTPLSPT-----PMFPFLPSAPSAPKLSHVTGIDTKDKSLK 503
DB 121 AASA--SVPAQOPPATTSATATPAANTTTGSPSAFPVQASRP-----NLVGSQLPPTTLP 174
QY 504 TVSSGAKGSELLSSESGALMEH---PEVSQVRKKTVEFN-----LTDMPPI--PENH 551
DB 175 VVSSNAQOQLPOOQLOOQOQOQPPQVSVAPLSNTAINGSPSTKETTTLPSVKAPEST 234
QY 552 LKEPLEQSPNTNIHTT 566
DB 235 LK---ETEPENNTS 246

RESULT 4

US-08-473-089-63
; Sequence 63, Application US/08473089
; Patent No. 6342368
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,089
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUPI, Fig. 46
US-08-473-089-63

Query Match 3.2%; Score 99.5; DB 4; Length 713;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
QY 349 AASAFRRASFMGSTFNISLNKEEMFQPNQDEE-----DAHAGI 389
DB 18 AIRQEFLOVSQEAITYRLQ-NOKDYDFKMNQQLAEMQOIRNTVYELELTHRKMKAAYEAE 76

QY 390 IGRF-IGLQSHDHPHPRANSRTKLLWPKRSLHLEGLPKNHKAQKONVRQEDNKAWK 448
DB 77 IKHLKLGLEQRDH-----QIASLTQV--QOQOQOQOQOQOHLQOQOQOLA 120
QY 449 AVDAFKSGPLYQOR-PCYYSAPOTPLSPT-----PMFPFLPSAPSAPKLSHVTGIDTKDKSLK 503
DB 121 AASA--SVPAQOPPATTSATATPAANTTTGSPSAFPVQASRP-----NLVGSQLPPTTLP 174
QY 504 TVSSGAKGSELLSSESGALMEH---PEVSQVRKKTVEFN-----LTDMPPI--PENH 551
DB 175 VVSSNAQOQLPOOQLOOQOQOQPPQVSVAPLSNTAINGSPSTKETTTLPSVKAPEST 234
QY 552 LKEPLEQSPNTNIHTT 566
DB 235 LK---ETEPENNTS 246

RESULT 5

US-08-487-072A-63
; Sequence 63, Application US/08487072A
; Patent No. 6423684
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,072A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUPI, Fig. 46
US-08-487-072A-63

Query Match 3.2%; Score 99.5; DB 4; Length 713;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
QY 349 AASAFRRASFMGSTFNISLNKEEMFQPNQDEE-----DAHAGI 389
DB 18 AIRQEFLOVSQEAITYRLQ-NOKDYDFKMNQQLAEMQOIRNTVYELELTHRKMKAAYEAE 76
QY 390 IGRF-IGLQSHDHPHPRANSRTKLLWPKRSLHLEGLPKNHKAQKONVRQEDNKAWK 448

Db 77 IKHLKLGLEQRDH-----QIASLTQV--QQQQQQQQQQVQOHLQQQQQOLA 120
 Qy 449 AVDAFKSGPLYQR-PGYISAPQTPLSPT-----PMFFPLEPSAPSKLSHVTGIDTKDKSLK 503
 Db 121 AASA--SVPAQOPPATTSATATPAANTTGTSPSAFFVQASRP-----NLVGSQLPPTTLP 174
 Qy 504 TVSSGAKKSFELLSESGALMEH---PEVSQVRKKTVEFN-----LTDMPET--PENH 551
 Db 175 VVSSNAQQQLPQOOLQOQQOQQOQQPPQVSVAPLSNTAINGSPTSKETTTLPSVKAPEST 234
 Qy 552 LKEPLEQSPNTIHTT 566
 Db 235 LK---ETEPENNTS 246

RESULT 6

US-09-108-857-3
 ; Sequence 3, Application US/09108857A
 ; Patent No. 6431137
 ; GENERAL INFORMATION:
 ; APPLICANT: JOHNSON, Alexander D.
 ; APPLICANT: BRAUN, Burkhard
 ; TITLE OF INVENTION: TUP1 SEQUENCES FROM CANDIDA ALBICANS AND METHODS FOR
 ; SCREENING AGENTS FOR INHIBITING VIRULENCE IN CANDIDA
 ; TITLE OF INVENTION: ALBICANS USING TUP1
 ; FILE REFERENCE: 22002-20006.00
 ; CURRENT APPLICATION NUMBER: US/09/108,857A
 ; CURRENT FILING DATE: 1998-07-01
 ; EARLIER APPLICATION NUMBER: 60/051,552
 ; EARLIER FILING DATE: 1997-07-02
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 713
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-09-108-857-3

Query Match 3.2%; Score 99.5; DB 4; Length 713;
 Best Local Similarity 23.1%; Pred. No. 0.43;
 Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
 Qy 349 AASAQFRASFQSGSTENISLNKEEMFQPNQDEE-----DAHAGI 389
 Db 18 ATRQEFLOVQSEANTRLQ-NQKDYDKMKNQQAEMQQTINTVYELTHRKNKDAYEAE 76
 Qy 390 IGRP-IGLQSHDHHPFRANSRTKLLWPQRESLLHGLPKNHKAQKQVRQEDNKAWKLK 448
 Db 77 IKHLKLGLEQRDH-----QIASLTQV--QQQQQQQQQQVQOHLQQQQQOLA 120
 Qy 449 AVDAFKSGPLYQR-PGYISAPQTPLSPT-----PMFFPLEPSAPSKLSHVTGIDTKDKSLK 503
 Db 121 AASA--SVPAQOPPATTSATATPAANTTGTSPSAFFVQASRP-----NLVGSQLPPTTLP 174
 Qy 504 TVSSGAKKSFELLSESGALMEH---PEVSQVRKKTVEFN-----LTDMPET--PENH 551
 Db 175 VVSSNAQQQLPQOOLQOQQOQQOQQPPQVSVAPLSNTAINGSPTSKETTTLPSVKAPEST 234
 Qy 552 LKEPLEQSPNTIHTT 566
 Db 235 LK---ETEPENNTS 246

RESULT 7

US-08-785-928-1
 ; Sequence 1, Application US/08785928
 ; Patent No. 6087115
 ; GENERAL INFORMATION:
 ; APPLICANT: Gershengorn, Marvin C.
 ; APPLICANT: Arvanitakis, Leandros
 ; APPLICANT: Geras-Raaka, Elizabeth
 ; APPLICANT: Cesarman, Ethel
 ; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE G PROTEIN COUPLED

; TITLE OF INVENTION: RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE
 ; TITLE OF INVENTION: ANTAGONISTS OF G PROTEIN COUPLED RECEPTORS
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 ; STREET: Clinton Square, P.O. Box 1051
 ; CITY: Rochester
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 14603
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/785,928
 ; FILING DATE: 22-JAN-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Goldman, Michael L.
 ; REGISTRATION NUMBER: 30,727
 ; REFERENCE/DOCKET NUMBER: 19603/1320
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (716) 263-1304
 ; TELEFAX: (716) 263-1600
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 342 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-785-928-1

Query Match 3.2%; Score 99; DB 3; Length 342;
 Best Local Similarity 25.0%; Pred. No. 0.15;
 Matches 70; Conservative 41; Mismatches 109; Indels 60; Gaps 15;
 Qy 3 IYTSQVANARLGSRLL--CWGSIYKL-LYGEFLIFLLCYIIIRFIRLALTEBQQ 59
 Db 68 VYIFCKRSRAGADIDLLGLCLNSLCISLSLAELMFLFPNIISTGLCRL-----E 121
 Qy 60 LMFEKILYCDISYIOLIPISFVLGPFYVTLVTVTRMNQYENLPWDRMLSLVSGFVEGKD- 118
 Db 122 IFYYLYVYLDIF-SVVCVSLV-RYLLVAYSTRSPKQSLGWLTSAALLIALVLSGA 179
 Qy 119 --EQSRL--LRRTLIIRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFWTPAEH 168
 Db 180 CRHSRVVDVPSKQAMCYENAGNMTADWRLHVRTVS-----VTAGFLLP--- 223
 Qy 169 KQLEKLSLPHMFWVPMVWTFANLSMKAWLGRIIRDPILL-----QSLLENMTL----- 217
 Db 224 --LALLILFYALTWCV-VRRTKLQARRKRGVIVAVLLFFVFCPPHYHVLNLLDTLRRR 280
 Qy 218 --RTQCGHLYAYDWISIPLYVT---QVTVVAVYSFFLTCL 252
 Db 281 WIRDSC---YTRGLINVLAVTSLQALYSVVPLIVSCL 317

RESULT 8

US-08-728-603-17
 ; Sequence 17, Application US/08728603
 ; Patent No. 6093806
 ; GENERAL INFORMATION:
 ; APPLICANT: Cesarman, Ethel
 ; APPLICANT: Knowles, Daniel M.
 ; TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED
 ; TITLE OF INVENTION: HERPESVIRUS
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
 ; STREET: Clinton Square, P.O. Box 1051

CITY: Rochester
 STATE: New York
 COUNTRY: USA
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/728,603
 FILING DATE: 10-OCT-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: BRAMAN, SUSAN J.
 REGISTRATION NUMBER: 34,103
 REFERENCE/DOCKET NUMBER: 19603/720
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 716-263-1636
 TELEFAX: 716-263-1600
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 342 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-728-603-17

Query Match 3.2%; Score 99; DB 3; Length 342;
 Best Local Similarity 25.0%; Pred. No. 0.15;
 Matches 70; Conservative 41; Mismatches 109; Indels 60; Gaps 15;
 QY 3 ITYTSQVANARLGSRLLL--CMRGSYKL-LYGEFLIFLCYIIRFIYRLALTEEQ 59
 DB 68 VTIYFKRHSRAGAITDILLGICLNSLCISLLAEVLMFLFNIISTGLCL-----E 121
 QY 60 LMPEKLTLYCSYIQLIPISFVLGVYVLVTRMNQYENLWPDRLMSLVSGFVEGKD- 118
 DB 122 IFYYLYVYLDIF-SVVCVSLV-RYLLVAYSTRSPKQSLGWLTSAALLIALVLSGD 179
 QY 119 --EQLR--LRRTLIRYANLGNV-----LILRSVSTAVYKRPSPSAQHLVQAGMTPAEH 168
 DB 180 CRHRSRVDPVSKQAMCTENAGMTADRLHVRTVS-----VTAGFLLP--- 223
 QY 169 KOLEKLSHPNMFVPMVPMFANLSMKAWLGGIRDPILL-----QSLNEMNTL--- 217
 DB 224 --LALLILFYALTWCV-VRRTKLQARRKVRGVI VAVVLLFFVFCFPYHVNLDTLLRRR 280
 QY 218 --RTQCHLYAYDWTISPLVYT---QVTVVAVYSFELTCL 252
 DB 281 WIRDSC---YTRGLINVLAVTSLILQALYSVAVPLIYSCL 317

RESULT 9
 US-08-722-255-6
 ; Sequence 6, Application US/08272255
 ; Patent No. 5824859
 ; GENERAL INFORMATION:
 ; APPLICANT: Cashmore, Anthony R.
 ; APPLICANT: Ahmad, Margaret
 ; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
 ; TITLE OF INVENTION: Using the same
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
 ; STREET: One Liberty Place, 46th floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/272,255
 FILING DATE: 08-JUL-1994
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Leary Ph.D., Kathryn
 REGISTRATION NUMBER: 36,317
 REFERENCE/DOCKET NUMBER: UPN-1795
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 681 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-272-255-6

Query Match 3.2%; Score 98.5; DB 2; Length 681;
 Best Local Similarity 18.9%; Pred. No. 0.5;
 Matches 93; Conservative 69; Mismatches 136; Indels 193; Gaps 27;
 QY 123 LLRRTLIRYANLGNV-----LILRSVSTAVYKRPSPSAQHLVQAGMTPAEHKQLEKLS 175
 DB 267 LVRIKQVAVANEGNAGESVNLFLKSLGIRYSYISFNH-----PYSH---ERPL 315
 QY 176 LPHNMFVPMVPMFANLSMKAWLGGIRDPILLQSLNEMNTLRTQCHLYAYDWTISPLV 235
 DB 316 LGHLKFF-PWVDENV-PKAWRQGRGTYP-LVDAGMRE-----LWATGMLH--- 358
 QY 236 YQVTVVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFF-----Y 284
 DB 359 --DRIRVVVSSFFVKV-----LQLPWRGMKYFWDTLDDADLESDA 397
 QY 285 VGWLKVA-----EOLINPFG-----DDDDPETNWI VDRNLQVSLAVDEMHDLP 330
 DB 398 LGWQVITGLPDSRFDRIDNPQEGYKPDNGEYVRRWLP-----ELSLPTDWIHP-- 451
 QY 331 RMEPDYMNKPEQPYPYTAASAQFRASPMGSTFNISLNKEBMEFQPNQEDBEDAHAGII 390
 DB 452 -----WNAPB--SVLQAAGIE-----LGSNYPLPI----- 474
 QY 391 GRFLGLQSHDHPHPRANSRTKLLWPKRESLLHGLPKNHKAQKQNVRGQEDNKAWKLKAV 450
 DB 475 ---VGLD-----EAKARLHEAL-----SOMWOLEAA 497
 QY 451 D--AFKSGPLYQRPQYYSAPQTLSPTMFFP-----LEPSAPSKLH-----SVT 493
 DB 498 SRAIENG---SEGLGDSAE--VEEAFIEFPDITMEETETRLNPNRREYEDQMVPISIT 552
 QY 494 GI-----DTKDKSLKTSSGAKKSPEL---LSESDGALME--HPEVSOVRRKTVFENLTD 543
 DB 553 SSLIRPEDEESSLNLNRSVGSRAEVPRNVMVNTQAQORRAEPASNOVTAMIPFNIRI 612
 QY 544 MPEIPENHLKE 554
 DB 613 VAESTEDSTAE 623

RESULT 10
 US-08-272-255-9
 ; Sequence 9, Application US/08272255
 ; Patent No. 5824859
 ; GENERAL INFORMATION:
 ; APPLICANT: Cashmore, Anthony R.
 ; APPLICANT: Ahmad, Margaret
 ; APPLICANT: Lin, Chentao
 ; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of

/ TITLE OF INVENTION: Using the Same
/ NUMBER OF SEQUENCES: 22
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
/ STREET: One Liberty Place, 46th floor
/ CITY: Philadelphia
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19103
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/272,255
/ FILING DATE: 08-JUL-1994
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Leary Ph.D., Kathryn
/ REGISTRATION NUMBER: 36,317
/ REFERENCE/DOCKET NUMBER: UPN-1795
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 568-3100
/ TELEFAX: (215) 568-3439
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 681 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-272-255-9

Query Match 3.2%; Score 98.5; DB 2; Length 681;
Best Local Similarity 18.9%; Pred. No. 0.5; Indels 193; Gaps 27;
Matches 93; Conservative 69; Mismatches 136;

QY	123	LLRRLTIRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLS	175
DB	267	LVRIRQVAVANEGNEAGESVNLFLKSGIGREYSRYSFNH-----PYSH---ERPL	315
QY	176	LPHNFWVFWVWFWANLSKAWLGGRIIDPILLOSLLENMTLRTQCGHLYAYDWTISIPLV	235
DB	316	LGHLLKFF-PWAVDENY-FKAWRQGRGTGYP-LVDAGMRE-----LWATGWLH----	358
QY	236	YQVVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOPFF-----Y	284
DB	359	--DRIRVVVSSFFVKV-----LQLPWRGWMKIFYDWTLLDADLESDA	397
QY	285	VGWLKVA-----EQLINPFG-----DDDFETNWIVDRLQVSLAVDEMHQDLPL	330
DB	398	LGWQVITGLPDSREFRIDNPQEGYKDPNGEVVRRMLP-----ELSLPTDWTIHP--	451
QY	331	RMEPDWYNKPEPPYTAASQAFRRASFGWSTFNISLNKEEMEPQNOEDEDAHAGII	390
DB	452	-----WNAPE--SVLQAAGIE-----LGSNYPLPI-----	474
QY	391	GRFLGLQSHDHHPPRANSRTKLLWPKRSLHLEGLPKNHKAQKQNVRGQEDNKAKLKAV	450
DB	475	---VGLD-----EAKARLHEAL-----SOMQLEAA	497
QY	451	D--AFKSGPLYQRPGYSAQPTLSTPMFFP-----LEPSAFSKLH-----SVT	493
DB	498	SRAALTEG---SEEGIGLSAE--VEEAPTEFFPRDITMEETETRLNPNRRYEDQMVPSIT	552
QY	494	GI-----DTKDKSLKTVSSGAKKSPEL-----LSESQALME--HPEVSQVRRKTVFNLTD	543
DB	553	SSLIPPEDEBSLNRNSVGSRAEVPVRNMTVNOAQORRAEPASQVMTAMIPFNTRI	612
QY	544	MPEIPENHLKE	554
DB	613	VAESTEDSTAE	623

RESULT 11
PCT-US95-08565-6
/ Sequence 6, Application PC/TUS9508565
/ GENERAL INFORMATION:
/ APPLICANT: Cashmore, Anthony R.
/ APPLICANT: Ahmad, Margaret
/ APPLICANT: Lin, Chentao
/ TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
/ TITLE OF INVENTION: Using the Same
/ NUMBER OF SEQUENCES: 22
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
/ STREET: One Liberty Place, 46th floor
/ CITY: Philadelphia
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19103
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/08565
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/272,255
/ FILING DATE: 08-JUL-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Leary Ph.D., Kathryn
/ REGISTRATION NUMBER: 36,317
/ REFERENCE/DOCKET NUMBER: UPN-1795
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 568-3100
/ TELEFAX: (215) 568-3439
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 681 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ PCT-US95-08565-6

Query Match 3.2%; Score 98.5; DB 5; Length 681;
Best Local Similarity 18.9%; Pred. No. 0.5; Indels 193; Gaps 27;
Matches 93; Conservative 69; Mismatches 136;

QY	123	LLRRLTIRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLS	175
DB	267	LVRIRQVAVANEGNEAGESVNLFLKSGIGREYSRYSFNH-----PYSH---ERPL	315
QY	176	LPHNFWVFWVWFWANLSKAWLGGRIIDPILLOSLLENMTLRTQCGHLYAYDWTISIPLV	235
DB	316	LGHLLKFF-PWAVDENY-FKAWRQGRGTGYP-LVDAGMRE-----LWATGWLH----	358
QY	236	YQVVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOPFF-----Y	284
DB	359	--DRIRVVVSSFFVKV-----LQLPWRGWMKIFYDWTLLDADLESDA	397
QY	285	VGWLKVA-----EQLINPFG-----DDDFETNWIVDRLQVSLAVDEMHQDLPL	330
DB	398	LGWQVITGLPDSREFRIDNPQEGYKDPNGEVVRRMLP-----ELSLPTDWTIHP--	451
QY	331	RMEPDWYNKPEPPYTAASQAFRRASFGWSTFNISLNKEEMEPQNOEDEDAHAGII	390
DB	452	-----WNAPE--SVLQAAGIE-----LGSNYPLPI-----	474
QY	391	GRFLGLQSHDHHPPRANSRTKLLWPKRSLHLEGLPKNHKAQKQNVRGQEDNKAKLKAV	450
DB	475	---VGLD-----EAKARLHEAL-----SOMQLEAA	497


```
; Patent No. 6465210
; GENERAL INFORMATION:
; APPLICANT: Peles, Elior
; TITLE OF INVENTION: CASPR/190, A FUNCTIONAL LIGAND FOR RTP-BETA AND THE
; TITLE OF INVENTION: AXONAL CELL RECOGNITION MOLECULE CONTACTIN
; FILE REFERENCE: 7683-111
; CURRENT APPLICATION NUMBER: US/08/826,134A
; CURRENT FILING DATE: 1997-03-26
; EARLIER APPLICATION NUMBER: 60/014,199
; EARLIER FILING DATE: 1996-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1384
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-08-826-134-2

Query Match      3.1%; Score 97.5; DB 4; Length 1384;
Best Local Similarity 20.5%; Pred. No. 2;
Matches 95; Conservative 50; Mismatches 180; Indels 139; Gaps 21;

QY 203 DPILLQSLNENMTLTQCHLYAYDWISPLVYTVV-----TVVYSFF---ITCLV 253
Db 725 DPALYCNCDADQPMRTDKGLLTFVDHLPV-----TQVIGDNNRSTSEAOFFLRPLRCYG 780
QY 254 GRQFLNPAKAYPG-----HELDLVVPVFTLQFFF-----YVG 286
Db 781 DRNSWNTISFHTGAALRPPPIRANHSLD-----VSFYFRTSAPSGVLENMGGPYCQ 832
QY 287 W-----LKVAQLINPF-----GDD-----DFETN---WIVDR---NLQVSLAYD 323
Db 833 WRRPVYRVELNTRSDVVPFADVGVNGDENLTVHSDDFENDEHVLVRAEINVKQARLVD 892
QY 324 EHQDLPRMEPDWYWKPPQPPYTAASQPRASFNGSTFNISLNKEEMEPQNOEDEE 383
Db 893 HRPWLRPMLQTYIMYEDQPLY-VGSAELKRRPVGCLRAMRLNGVTL-----NLEGRA 947
QY 384 DAHAGIIGRFLGLQSHDH-----HPPRANSRTKLLWPKRESLLHGLPKNHKAQKQNVRGQ 439
Db 948 NASEGTSNCTGHCAHPLPCFHGRCVERYSYTTCCDLTAFDGPYCNH-----DIGGF 1002
QY 440 ENKAWK-----LKAVDAPKSG--PLYORPGY-----YSAPOT 470
Db 1003 PEPGTWMYNLOSALRSAAAREFSHMLSRVPVGPYGPYIPGYDTPGVPGYHGPYRLPDY 1062
QY 471 PLSPMPFPPLPSAPSCLKHSVTG-----IDTKDKSLKTVSSGAKKSFELLSQDCA 522
Db 1063 PRPGRPV-----PGYRGPVYNTGERSVFSSTSSAPAVLLYVSSFVDRYMAVLKDDGT 1117
QY 523 LMEHPVSQVRRTKVEFNLTDMPEIPENHLKPELQSPNTIHTT 566
Db 1118 LQ-----LRYQLGTSPPYVQLTTRPVTGQPHSINIT 1149

RESULT 15
US-08-885-291-55
; Sequence 55, Application US/08885291A
; Patent No. 6057125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885,291A
; CURRENT FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 08/816,693
; EARLIER FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 846
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-885-291-55

Query Match      3.1%; Score 97; DB 3; Length 846;
Best Local Similarity 25.6%; Pred. No. 1;
Matches 45; Conservative 28; Mismatches 57; Indels 46; Gaps 10;

QY 416 KEESLLHEGLPKNHKAQKQNVGQEDNKAWKL---KAVDAFKSGPL-----YORPGYYS 466
Db 386 RRELGIIESLPET---AADKSQDSGSDNRINTVSLKEALERFDHSPPTPSASSRSRKSST 443
QY 467 APQTPLSPPTMPFPLEPSAPSKLHVTGIDTKDKSLKTVSSGAKKSFELLSQDQALMEH 526
Db 444 AVSDP-SSTPTKIPTDTSTPPROH-----LPAHEKQVQVRSFSSQIN--SQVGSLLTQ 496
QY 527 PEVSQVRRTKVEFNLTDMPEIPE-----NHLKEPLEQS-----PTNIH 564
Db 497 PYMSQA-----TNLP-IPQGMSSQFQSAQLGAMQHLKQLEQRTRMIEANIH 542

Search completed: July 10, 2003, 12:27:52
Job time : 16.4055 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:19:13 ; Search time 36.1203 Seconds
(without alignments)
2158.111 Million cell updates/sec

Title: US-09-622-964-3

Perfect score: 3120

Sequence: 1 MTITYTQVANRLGFSRLL.....TLKDHMPYALENDEAHS 585

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.++

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3120	100.0	585	20	Human CG1CE long f
2	1899	60.9	551	20	Mouse CG1CE protei
3	1535	49.2	435	20	Human CG1CE short
4	1397	44.8	261	20	Human adult retina
5	1009	32.3	721	22	Drosophila melanog
6	886	28.4	535	22	Drosophila melanog
7	824.5	26.4	368	22	Novel human diagno
8	742	23.8	327	22	Drosophila melanog
9	694.5	22.3	731	22	Drosophila melanog
10	661	21.2	314	22	Human secreted pro

11	549	17.6	188	23	AB889288	Human polypeptide
12	539	17.3	251	22	AA870067	Human secreted pro
13	539	17.3	251	23	ABG65506	Human albumin fusi
14	523	16.8	101	22	ABB14768	Human nervous syst
15	324.5	10.4	701	22	ABG08608	Novel human diagno
16	238	7.6	99	22	AAU22652	Novel human colon
17	238	7.6	99	22	AAU22652	Human digestive sy
18	205.5	6.6	377	22	ABG06411	Novel human diagno
19	174	5.6	30	20	AAU29956	Human CG1CE partia
20	126	4.0	47	22	ABG08607	Novel human diagno
21	118	3.8	30	20	AAU29971	C. elegans CG1CE r
22	117	3.8	30	20	AAU29968	C. elegans CG1CE r
23	115	3.7	30	20	AAU29969	C. elegans CG1CE r
24	112	3.6	1256	22	ABG60532	Drosophila melanog
25	111.5	3.6	637	22	AAU23162	Novel human enzyme
26	111.5	3.6	637	22	AAU23701	Novel human enzyme
27	111.5	3.6	832	22	AAO13516	Human polypeptide
28	111.5	3.6	849	19	AAW82397	Human UBP protein
29	109	3.5	283	21	AA42835	Human ORFX ORF2599
30	109	3.5	1051	21	AA42835	Arabidopsis thalia
31	109	3.5	1057	21	AA42835	Arabidopsis thalia
32	107	3.4	384	21	AA26452	Drosophila melanog
33	107	3.4	384	22	ABG6539	Drosophila melanog
34	107	3.4	389	21	ABG20926	Drosophila odorant
35	107	3.4	722	22	ABG65012	Drosophila melanog
36	106	3.4	581	18	AAW22103	Murine leptin rece
37	105	3.4	518	22	AAW39378	Human polypeptide
38	105	3.4	531	22	AA41164	Human polypeptide
39	105	3.4	1816	21	AAU95440	Caenorhabditis ele
40	104.5	3.3	785	23	ABG1027	Herbicidally activ
41	104.5	3.3	786	22	AAU34510	E. coli cellular p
42	103.5	3.3	356	22	AAU93521	Human protein sequ
43	103.5	3.3	356	23	ABG97330	Novel human protei
44	103	3.3	30	20	AAU29970	C. elegans CG1CE r
45	103	3.3	854	21	AA42329	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAU29953
ID AAU29953 standard; Protein; 585 AA.
XX
AC AAU29953;
XX
XX
DT 22-NOV-1999 (first entry)
XX
DE Human CG1CE long form protein sequence.
XX
KW CG1CE; Best's macular dystrophy; mutation; diagnosis; detection;
KW BMD; age-related macular dystrophy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Misc-difference 121
FT /note= "encoded by GGC"
XX
XX WO9943695-A1.
XX
XX 02-SEP-1999.
XX
XX 22-FEB-1999; 99WO-US03790.
XX
XX 25-FEB-1998; 98US-0075941.
XX
XX 18-DEC-1998; 98US-0112926.
XX
XX (MERI) MERCK & CO INC.
XX
XX (UYUP-) UNIV UPPSALA.
XX
XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;

DR WPI; 1999-540560/45.
 XX N-PSDB; AA221227.
 PT Human and mouse polynucleotides encoding CG1CE polypeptides -
 XX Claim 7; Fig 3; 67pp; English.
 XX The present sequence represents the human CG1CE protein. When the CG1CE
 CC gene is mutated it is responsible for Best's macular dystrophy (BMD).
 CC Polynucleotides encoding CG1CE are useful for diagnosing whether a
 CC patient carries a mutation in the CG1CE gene. Normal and mutated
 CC CG1CE proteins are useful for identifying activators and/or inhibitors
 CC of these proteins, in order to treat BMD. The CG1CE gene offers a
 CC simpler and cheaper method of diagnosing BMD without the need for the
 CC presence of the patient. The gene may also be useful to discovering
 CC the genetic cause of age-related macular dystrophy.
 XX Sequence 585 AA;
 SQ Query Match 100.0%; Score 3120; DB 20; Length 585;
 Best Local Similarity 100.0%; Pred. No. 5e-296;
 Matches 585; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
 Db 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
 QY 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 Db 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 QY 121 SLLLRRTLRANLGNVLIRSVSTAVYKRFPSAHLVQAGFMTPAEHKQLEKLSLPHNM 180
 Db 121 SLLLRRTLRANLGNVLIRSVSTAVYKRFPSAHLVQAGFMTPAEHKQLEKLSLPHNM 180
 QY 181 FWVPWFWFANLGNLGRIRDPILLOSLLNEMTLRTCCGHLAYDWISIPLVTVQV 240
 Db 181 FWVPWFWFANLGNLGRIRDPILLOSLLNEMTLRTCCGHLAYDWISIPLVTVQV 240
 QY 241 TVAVTSFPLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 Db 241 TVAVTSFPLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 QY 301 DDDDFETNWI VDRNLQVSLA VDEMHDLP RMEPDMYWNKPEPPYTAASAQFRASFM 360
 Db 301 DDDDFETNWI VDRNLQVSLA VDEMHDLP RMEPDMYWNKPEPPYTAASAQFRASFM 360
 QY 361 GSTFNISLNKEEMEFQPNQDEDAHAGIIGRFLGQSHDHPHPRANSRTKLLWPK 420
 Db 361 GSTFNISLNKEEMEFQPNQDEDAHAGIIGRFLGQSHDHPHPRANSRTKLLWPK 420
 QY 421 LHEGLPKNHKAQKQNVRCQEDNKAWLKAVDAFKSGPLVYQRPYYSAPOTPLSPTPMFPF 480
 Db 421 LHEGLPKNHKAQKQNVRCQEDNKAWLKAVDAFKSGPLVYQRPYYSAPOTPLSPTPMFPF 480
 QY 481 LEPSPAKSLHVTGIDTKDKSLKTVSSGAKSFELLSESDGALMEHPEVSQVRKRTVEFN 540
 Db 481 LEPSPAKSLHVTGIDTKDKSLKTVSSGAKSFELLSESDGALMEHPEVSQVRKRTVEFN 540
 QY 541 LTDMPPEIPENHLKEPLEQSPNTIHTTLKDHMDPYWALENREDAHS 585
 Db 541 LTDMPPEIPENHLKEPLEQSPNTIHTTLKDHMDPYWALENREDAHS 585
 RESULT 2
 ID AAY29955 standard; Protein; 551 AA.
 AC AAY29955;
 XX AAY29955;
 XX 22-NOV-1999 (first entry)
 XX Mouse CG1CE protein sequence.

XX CG1CE; Best's macular dystrophy; mutation; diagnosis; detection;
 KW BMD; age-related macular dystrophy.
 XX Mus sp.
 XX WO9943695-A1.
 XX 02-SRP-1999.
 XX 22-FEB-1999; 99WO-US03790.
 XX 25-FEB-1998; 98US-0075941.
 XX 18-DEC-1998; 98US-0112926.
 XX (MERI) MERCK & CO INC.
 XX (UYUP-) UNIV UPPSALA.
 XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;
 DR WPI; 1999-540560/45.
 DR N-PSDB; AA221229.
 XX Human and mouse polynucleotides encoding CG1CE polypeptides -
 PT Claim 7; Fig 8; 67pp; English.
 XX The present sequence represents the mouse CG1CE protein. When the CG1CE
 CC gene is mutated it is responsible for Best's macular dystrophy (BMD).
 CC Polynucleotides encoding CG1CE are useful for diagnosing whether a
 CC patient carries a mutation in the CG1CE gene. Normal and mutated
 CC CG1CE proteins are useful for identifying activators and/or inhibitors
 CC of these proteins, in order to treat BMD. The CG1CE gene offers a
 CC simpler and cheaper method of diagnosing BMD without the need for the
 CC presence of the patient. The gene may also be useful to discovering
 CC the genetic cause of age-related macular dystrophy.
 XX Sequence 551 AA;
 SQ Query Match 60.9%; Score 1899; DB 20; Length 551;
 Best Local Similarity 64.3%; Pred. No. 1.3e-176;
 Matches 378; Conservative 59; Mismatches 101; Indels 50; Gaps 10;
 QY 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
 Db 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
 QY 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 Db 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 QY 121 SLLLRRTLRANLGNVLIRSVSTAVYKRFPSAHLVQAGFMTPAEHKQLEKLSLPHNM 180
 Db 121 SLLLRRTLRANLGNVLIRSVSTAVYKRFPSAHLVQAGFMTPAEHKQLEKLSLPHNM 180
 QY 181 FWVPWFWFANLGNLGRIRDPILLOSLLNEMTLRTCCGHLAYDWISIPLVTVQV 240
 Db 181 FWVPWFWFANLGNLGRIRDPILLOSLLNEMTLRTCCGHLAYDWISIPLVTVQV 240
 QY 241 TVAVTSFPLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 Db 241 TVAVTSFPLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 QY 301 DDDDFETNWI VDRNLQVSLA VDEMHDLP RMEPDMYWNKPEPPYTAASAQFRASFM 360
 Db 301 DDDDFETNWI VDRNLQVSLA VDEMHDLP RMEPDMYWNKPEPPYTAASAQFRASFM 360
 QY 361 GSTFNISLNKEEMEFQPNQDEDAHAGIIGRFLGQSHDHPHPRANSRTKLLWPK 416
 Db 361 GSTFNISLNKEEMEFQPNQDEDAHAGIIGRFLGQSHDHPHPRANSRTKLLWPK 416
 QY 417 RESLLHEGLPKNHKAQKQNVRCQEDNKAWLKAVDAFKSGPLVYQRPYYSAPOTPLSPTP 476
 Db 417 RESLLHEGLPKNHKAQKQNVRCQEDNKAWLKAVDAFKSGPLVYQRPYYSAPOTPLSPTP 476

Db 421 NPLL--BQCKD-----ANQNKQD--VWFKGLDFLKVPRFKRSGHCGQAPSS--- 468

Qy 477 MFPPLPSASKLSVGTGIDTKSLKTVSSGAKKSPILLSGALMEHPEVSQVRRKT 536

Db 469 --HPTQOSAPSS--SDTG-----DGPSTDYQIECHMKKKT 499

Qy 537 VEFNLTDMPETPENHKE-PLEQSPNTHTTLKDHMPYWALENRDEA 583

Db 500 VEFNL-NIPESPTEHLQORLDQMSNTIQLMKHEAESY---PYRDEA 543

RESULT 3

AAV29954

ID AAY29954 standard; Protein; 435 AA.

XX AC AAY29954;

XX DT 22-NOV-1999 (first entry)

XX DE Human CGICE short form protein sequence.

XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

XX KW BMD; age-related macular dystrophy.

XX OS Homo sapiens.

XX PN WO9943695-A1.

XX PD 02-SEP-1999.

XX PF 22-FEB-1999; 99WO-US03790.

XX PR 25-FEB-1998; 98US-0075941.

XX PR 18-DEC-1998; 98US-0112926.

XX PA (MERI) MERCK & CO INC.

XX PA (UYUP-) UNIV UPPSALA.

XX PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;

XX DR WPI; 1999-540560/45.

XX DR N-PSDB; AA221228.

XX Human and mouse polynucleotides encoding CGICE polypeptides

PS Claim 7; Fig 5; 67pp; English.

CC The present sequence represents the human CGICE protein. When the CGICE gene is mutated it is responsible for Best's macular dystrophy (BMD). Polynucleotides encoding CGICE are useful for diagnosing whether a patient carries a mutation in the CGICE gene. Normal and mutated CGICE proteins are useful for identifying activators and/or inhibitors of these proteins, in order to treat BMD. The CGICE gene offers a simpler and cheaper method of diagnosing BMD without the need for the presence of the patient. The gene may also be useful to discovering the genetic cause of age-related macular dystrophy.

SQ Sequence 435 AA;

Query Match 49.2%; Score 1535; DB 20; Length 435;

Best Local Similarity 99.7%; Pred. No. 3.7e-141;

Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

Db 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

Qy 61 MFEKLTLYCDSYIQLIPISFVLGPFVTVVVTWNNQYENLPDPRLMSLVSGFVGKDEQ 120

Db 61 MFEKLTLYCDSYIQLIPISFVLGPFVTVVVTWNNQYENLPDPRLMSLVSGFVGKDEQ 120

Qy 121 SRLLRRTLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

Db 121 GLLRRTLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

Qy 181 FWVPVWVFANLSWKAMLGGRIRDPILLOSILNENMTLRTQCHLYAYDWISIPLVYQTV 240

Db 181 FWVPVWVFANLSWKAMLGGRIRDPILLOSILNENMTLRTQCHLYAYDWISIPLVYQTV 240

Qy 241 TVAVYSPFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFFVYVGLKV 290

Db 241 TVAVYSPFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFFVYVGLKV 290

RESULT 4

AAW95345

ID AAW95345 standard; Protein; 261 AA.

XX AC AAW95345;

XX DT 26-APR-1999 (first entry)

XX DE Human adult retina secreted protein bk112_15.

XX KW Secreted protein; human; retina; bk112_15.

XX OS Homo sapiens.

XX PN WO9856909-A2.

XX PD 17-DEC-1998.

XX PF 08-JUN-1998; 98WO-US11822.

XX PR 05-JUN-1998; 98US-0092722.

XX PR 11-JUN-1997; 97US-0873218.

XX PA (GEMY) GENETICS INST INC.

XX PI Agostino MJ, Fichtel K, Howes SH, Jacobs K, LaVallie ER;

XX PI McCoy JM, Racie LA, Spaulding V, Treacy M;

XX DR WPI; 1999-080899/07.

XX DR N-PSDB; AAV99722.

XX New polynucleotides encoding secreted human proteins - derived from human foetal brain, adult testes, foetal kidney, adult thyroid or adult retina cDNA libraries

PS Claim 13; Page 72-73; 113pp; English.

CC This is the amino acid sequence of bk112_15, a novel human secreted protein predicted from the nucleotide sequence of a human adult retina cDNA clone (see AAV99722). Database searches indicate some sequence similarity to known sequences. The invention provides cDNA clones (see AAV99721-33) from human adult thyroid, adult retina, adult testis, foetal kidney and foetal brain that encode novel secreted proteins (see AAW95344-53). The polynucleotides and proteins are predicted to have activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional, cytokine, cell proliferation/differentiation, immune stimulating (e.g. as vaccines) or immune suppressing, haematopoiesis regulating, tissue growth, activin/inhibin, chemotactic/chemokinetic, haemostatic, thrombolytic, receptor/ligand, antiinflammatory, cadherin/tumour invasion suppressor, and tumour inhibition activities.

XX SQ Sequence 261 AA;

Query Match 44.8%; Score 1397; DB 20; Length 261;

Best Local Similarity 99.6%; Pred. No. 5.4e-128;

Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 325 MHQDLPRMEPDMYWNKPQPPYTAASAQFRASFMGSTFNISLNKEEMFQNOEDEED 384

Db 1 MHQDLPRMEPDYWNKPEOPPYTAASQAQFRASFMGSTFNISLNKEMEFPQNOEDE 60
 QY 385 AHAGIIGRFLGLOSHDHPHPRANSRTKLLWPKRESLLHGLPKNHKAQONVRGOENKA 444
 Db 61 AHAGIIGRFLGLOSHDHPHPRANSRTKLLWPKRESLLHGLPKNHKAQONVRGOENKA 120
 QY 445 WKUKAVDAFKSGPLYQRPYYSAPQTPLSPTMFPFPLEPSAPSXLHVSVTGDTKDKSLKT 504
 Db 121 WKUKAVDAFKSAPLYQRPYYSAPQTPLSPTMFPFPLEPSAPSXLHVSVTGDTKDKSLKT 180
 QY 505 VSSGAKKSFELSSDGLMEHPVSVQVRKTVFNLTDMPETIPENHLKEPLEOSPNNIH 564
 Db 181 VSSGAKKSFELSSDGLMEHPVSVQVRKTVFNLTDMPETIPENHLKEPLEOSPNNIH 240
 QY 565 TTLKDHMDPYWALENDRDEAHS 585
 Db 241 TTLKDHMDPYWALENDRDEAHS 261

RESULT 5 ABB62213

ID ABB62213 standard; Protein; 721 AA.

AC ABB62213;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 13431.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL06316.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions

XX Disclosure; SEQ ID NO 13431; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 721 AA;

XX Query Match

XX Best Local Similarity 32.3%; Score 1009; DB 22; Length 721;

XX 36.7%; Pred. No. 2.2e-89;

Matches 250; Conservative 83; Mismatches 216; Indels 132; Gaps 18;

QY 1 MTITYTSOANAR-LGSFRLLLCWGSIYKLLYGEFLFLCYIIRFYRLALTEEQ 59

Db 1 MTITYTGEVATCGFCFLKLLLRWRSYKLVLDLLAPLTIYIYAINMYRYFGLNPAQK 60

QY 60 LMFPEKLTLYCDSYIOLIPISFVLGFTVTLVTRWNNQYENLPWDRMLSVSGFVEGKDE 119

Db 61 ETEFAIVQYCDSEIRELIPISFVLGFTVTRWNNQYISIPWDPFIADVSSNVHGQDE 120

QY 120 QSLRLRTLRVANGLNVLILRSVTAIVKRPFSQAHLVQAGFMTFAEHKQLEKL--SLP 177

Db 121 RGMERTIMRYVCLCTWLANVSPVKRPFGLNNLVEAGLLDNKEKTIITMKNKAP 180

QY 178 H-NMFVWVWVWFAANLWGLGRIRDPILLQSLNEMNTLRQCCHLYAYDHSIPLVY 236

Db 181 RPSKHWLPVWAAISIIRAKRGRIHDDFAVKTIIDELNKRQCGLLISYDTISVPLVY 240

QY 237 TQVTVVAVYSFFLTCLVGRQ-----FLNPAKAYPGHELDLVVPVFTLQEFFYV 285

Db 241 TQVTVLAVYSFLTCCMGQQTGKVVGNVTYLN-----KVDLYFPVFTTLQEFFYV 292

QY 286 GMLKVAEQLINPFGEDDDDPETNWIVDRLQVSLAVDENHQDLPRMEPDYMNKPEP-Q 344

Db 293 GMLKVAESLINPFGEDDDDFEVNMVYDRNLQVSYLIVDEMHHDHPELLKQYWDEVFPNE 352

QY 345 PVTAAQAQFRASFMGSTFNISLNK-----EEMEFQPNQOEDEDAH----- 386

Db 353 LPTTAAERFRENHPEPSTAKIEVPKNAAMPSTMSVRIDEMADDASGIFHSAGNGKML 412

QY 387 -----AGIIGRPLG-----LQSHDHHPHPRANSRTKLLWPK 416

Db 413 DSSPSLVSVSGTLSRVNTVASALKRFLSRDDSRPGSATSPQOPYKFPASASASLSGAV 472

QY 417 RESLHLEGLPKNHKAQONVRGOENKA--WKLKAVD-----AFKSGPLY 459

Db 473 VGSATSAGKPGASLRITQOVIIEVDEQATITSMRANDPRPNVMDIPAQTSAGTSGPLQ 532

QY 460 QRPGYYSAPQTPLSP-----TPMFFP-----LEPSAPS-----KLHVSVTG-- 494

Db 533 PPAHSEPVDIRPSPSYNRAOSQVEPNLFPGGVDALLSTAPAGGSPLLLSNAATAPS 592

QY 495 --IDTKDKSLKTVSSGAKK---SPELISESD--GALMEHPVSVQVRKTVFNLTDMP 546

Db 593 SPVGSSKSLYDPQKASRETVESMDLSRSTLLGDAAVQPEDEG-----DDFDK 642

QY 547 IPENHLKEPLEOSPNNIH 567

Db 643 LKAEREKEKLMRQKNLARTI 663

RESULT 6

ABB63121

ID ABB63121 standard; Protein; 535 AA.

XX AC ABB63121;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 16155.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PD 29-NOV-2001.
 XX 18-MAY-2001; 2001WO-US16450.
 XX 19-MAY-2000; 2000US-205515P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Birse CE, Rosen CA;
 XX WPI; 2002-122018/16.
 XX N-PSDB; ABL89697.
 DR Novel 1405 isolated polypeptides, useful for diagnosis, treatment, and
 XX prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 XX Claim 11; SEQ ID NO 1664; 2081pp + Sequence Listing; English.
 PS The invention relates to novel genes (ABL99449-ABL90853) and proteins
 XX (ABB89040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 188 AA;
 Query Match 17.6%; Score 549; DB 23; Length 188;
 Best Local Similarity 97.2%; Pred. No. 3e-45;
 Matches 105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 477 MFPPLEPSAPSKLHVSVTGIDTKSLKTVSSGAKKSFELLSESDGALMEHPEVSVRRKT 536
 Db 1 MFPPLEPSAPSKLHVSVTGIDTKSLKTVSSGAKKSFELLSESDGALMEHPEVSVRRKT 60
 QY 537 VFENLTDMPETPENHLKEPLEQSPPTNIHTTLKDHMDPYWALENRDEAH 584
 Db 61 VFENLTDMPETPENHLKEPLEQSPPTNIHTTLKDHMDPYWALENRSLVH 108
 RESULT 12
 AAB70067
 ID AAB70067 standard; Protein; 251 AA.
 XX AAB70067;
 AC
 DT 14-MAY-2001 (first entry)
 XX Human secreted protein #6.
 DE Human; secreted protein; immunomodulatory; anti-sclerotic;
 KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
 KW immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
 KW antialzheimer's; antiparkinsonian; antimicrobial; vulnary; gene therapy;
 KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
 KW neurological; infection.
 XX Homo sapiens.
 OS

XX WO200112776-A2.
 XX 22-FEB-2001.
 XX 15-AUG-2000; 2000WO-US22350.
 XX 16-AUG-1999; 99US-0148759.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;
 XX WPI; 2001-244245/25.
 DR N-PSDB; AAF76848.
 XX Nucleic acids encoding 18 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX Claim 11; Page 362-363; 380pp; English.
 PS The present sequence is one of 18 novel human secreted proteins. The
 XX nucleic acids encoding the proteins and the proteins themselves may be
 CC used in the prevention, diagnosis and treatment of diseases including
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
 CC and human immunodeficiency virus (HIV) infections), hyperproliferative
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
 CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC infectious diseases and/or for promoting wound healing, regeneration
 CC and/or chemotaxis. The nucleic acid molecules may also be used as DNA
 CC probes in diagnostic assays to detect and quantitate the presence of
 CC similar nucleic acid sequences in samples. The polypeptides may also be
 CC used as antigens in the production of antibodies and in assays to
 CC identify modulators of protein expression and activity.
 XX SQ Sequence 251 AA;
 Query Match 17.3%; Score 539; DB 22; Length 251;
 Best Local Similarity 55.5%; Pred. No. 4.4e-44;
 Matches 111; Conservative 26; Mismatches 31; Indels 32; Gaps 6;
 QY 224 LYAYDWISIPLVYTVQVTVAVYSFFLTCLVGRQLNP-----AKAY-----PGHE----- 268
 Db 2 LFHYDWISIPLVYTVQVTVAVYSFFLTCLVGRQLNP-----AKAY-----PGHE----- 268
 QY 269 LDLVVPVFTFLQFFVYVGMVKVAQLINPFGDDDDDDFTNWIIVDRNLQVSLLAIVDEMHD 328
 Db 62 PMYVPLATLLQFFVYVGMVKVAQLINPFGDDDDDDFTNWIIVDRNLQVSLLAIVDEMHD 121
 QY 329 LPRMEPDWYMKPEPQPPYT-AASQFRASFWGSTFNISLNKEEFQPNQDEDAHA 387
 Db 122 LPPAEKQYWDQDQPPYTVATAESLRFSLGTFNLRMSDD-----PEQSLQVEASP 176
 QY 388 G-----IIIGRFLGL 396
 Db 177 GSGRPAPAAQTPLGLGLV 196
 RESULT 13
 ABG65506
 ID ABG65506 standard; Protein; 251 AA.
 XX ABG65506;
 AC ABG65506;
 XX 27-AUG-2002 (first entry)
 DT Human albumin fusion protein #2181.
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW

KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antifertility; antiinflammatory; antitumor;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.

OS Homo sapiens.
 OS Synthetic.

PN WO20017137-A1.

XX 18-OCT-2001.

XX 12-APR-2001; 2001WO-US11988.

XX 12-APR-2000; 2000US-229358P.

PR 25-APR-2000; 2000US-199384P.

PR 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Hasseltine WA;

XX WPI; 2002-010886/01.

DR New fusion protein for treating disease e.g. diabetes comprises an

XX albumin fused to a therapeutic protein.

PT Claim 1; Page 2075; 2102pp; English.

XX The present invention relates to albumin fusion proteins comprising a

CC therapeutic protein X and human albumin (HA, also known as human serum

CC albumin, HSA). The proteins are useful for treating a disease or

CC disorder that may be modulated by therapeutic protein X. The albumin

CC extends the shelf-life of protein X, and may increase its biological

CC in vitro/in vivo activity. The protein is useful for treating and

CC diagnosing disorders such as cancer, reproductive disorders, digestive

CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders

CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders

CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,

CC encephalomyelitis, meningitis, schizophrenia), and connective disorders

CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin

CC fusion proteins of the invention.

XX SQ Sequence 251 AA;

Query Match / 17.3%; Score 539; DB 23; Length 251;

Best Local Similarity 55.5%; Pred. No. 4.4e-44;

Matches 111; Conservative 26; Mismatches 31; Indels 32; Gaps 6;

QY 224 LYADWISIPVYVTVVAVYSFPLTCLVGRQELNP-----AKAY-----PGHE----- 268

Db 2 LFHYDWISIPVYVTVVAVYSFPLTCLVGRQELNP-----AKAY-----PGHE----- 61

QY 269 LDVVVFVTFQFFYVGVKVAEQIINPFGEDDDDDFETNIVDRNLQVSLVAVDEMHDQ 328

Db 62 PDVYVPLATLLOFFFYAGWLVKVAEQIINPFGEDDDDDFETNQLIDRNLQVSLVAVDEMVD 121

QY 329 LPRMEPDVYVTFQFFYVGVKVAEQIINPFGEDDDDDFETNIVDRNLQVSLVAVDEMHDQ 387

Db 122 LPPAEKQYWDQDQPPYVTVVAVYSFPLTCLVGRQELNP-----AKAY-----PGHE----- 176

QY 388 G-----IIGRFLGL 396

Db 177 GSGRPAPAAQPTLGRFLGV 196

RESULT 14

ABBI4768

ID ABB14768 standard; Protein; 101 AA.

XX AC ABB14768;
 XX DT 23-JAN-2002 (first entry)
 XX DE Human nervous system related polypeptide SEQ ID NO 3425.
 XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnary;
 KW antiparkinsonian; antitickling; antianaemic; antiarthritic; cancer;
 KW antineuritic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
 XX OS Homo sapiens.
 XX WO200159063-A2.
 XX 16-AUG-2001.
 XX 17-JAN-2001; 2001WO-US01334.
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-541565/60.
 DR N-PSDB; ABA11094.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 useful for preventing, diagnosing and/or treating nervous system
 cancers and metastases -
 PT
 XX
 PS Claim 11; SEQ ID NO 3425; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AB11004-ABA21534) and proteins
 (AB14678-AB18001) useful for preventing, treating or ameliorating
 medical conditions e.g. by protein or gene therapy. The genes are
 isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 and ovarian cancer and other cancers of the adrenal gland, bone, bone
 marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 printed specification, but was obtained in electronic format directly
 from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 101 AA;
 Query Match 16.8%; Score 523; DB 22; Length 101;
 Best Local Similarity 99.0%; Pred. No. 4.2e-43;
 Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 291 AEQLINPFGEDDDPETNWIYDRNLQVSLAVDENHQLPRMEPDYMNKPPQPYTAA 350
 DB 1 AEQLINPFGEDDDPETNWIYDRNLQVSLAVDENHQLPRMEPDYMNKPPQPYTAA 60
 QY 351 SAQFRASFMGSTFNISLNKEEMEFQPNQDEEDAHAG 388
 DB 61 SAQFRASFMGSTFNISLNKEEMEFQPNQDEEDAHAG 98
 RESULT 15
 ABC08608
 ID ABG08608 standard; Protein; 701 AA.
 XX
 AC ABG08608;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #8599.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:20:43 ; Search time 9.68173 Seconds
(without alignments)
2506.127 Million cell updates/sec

Title: US-09-622-964-3

Perfect score: 3120

Sequence: 1 MITYTSQVANRLGFSRL.....TLKHMDPYWALENDEAHS 585

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3110	99.7	585	1 VMD2 HUMAN	O76090 homo sapien
2	752	24.1	632	1 YQ22 CAEEL	P34672 caenorhabdi
3	751	24.1	405	1 YQ32 CAEEL	O09379 caenorhabdi
4	717	23.0	499	1 YHDI CAEEL	Q23369 caenorhabdi
5	682	21.9	413	1 YV4Q CAEEL	O45435 caenorhabdi
6	679	21.8	450	1 YQ23 CAEEL	Q17528 caenorhabdi
7	669.5	21.5	523	1 YQ24 CAEEL	Q17529 caenorhabdi
8	656	21.0	459	1 YK78 CAEEL	P34319 caenorhabdi
9	646.5	20.7	513	1 YXAK CAEEL	Q21373 caenorhabdi
10	636	20.4	456	1 YN4 CAEEL	P34577 caenorhabdi
11	622.5	20.0	501	1 YSV1 CAEEL	Q22566 caenorhabdi
12	622	19.9	420	1 YCEL CAEEL	Q94175 caenorhabdi
13	585	18.8	400	1 YV6L CAEEL	Q19978 caenorhabdi
14	503.5	16.1	884	1 YAVK CAEEL	Q17851 caenorhabdi
15	474.5	15.2	602	1 YH34 CAEEL	O18303 caenorhabdi
16	441.5	14.2	411	1 YH55 CAEEL	O45363 caenorhabdi
17	395	12.7	434	1 YSWJ CAEEL	O45363 caenorhabdi
18	390	12.5	81	1 VMD2 MOUSE	O88870 mus musculus
19	127.5	4.1	315	1 YNEE SALTI	O82706 salmonella
20	127.5	4.1	315	1 YNEE SALTY	Q82706 salmonella
21	115.5	3.7	613	1 YBID VIBCH	Q9kv48 vibrio chol
22	113	3.6	932	1 YAJA SCHPO	Q09897 schizosacch
23	111.5	3.6	798	1 YBPA HUMAN	Q14694 homo sapien
24	107	3.4	389	1 O85C DROME	Q9vbn6 drosophila
25	107	3.4	961	1 YF02 YEAST	P54860 saccharomyc
26	105	3.4	518	1 ATEL HUMAN	O95260 homo sapien
27	104.5	3.3	457	1 SECT CHLTR	P78539 chlamydia t
28	104.5	3.3	741	1 YBIO ECOLI	P75873 escherichia
29	104.5	3.3	1215	1 YBGN DROME	Q9w112 drosophila
30	103	3.3	321	1 YNEE ECO57	Q8xaz3 escherichia
31	102.5	3.3	321	1 YNEE ECOLI	P76146 escherichia
32	101.5	3.3	1385	1 CTA1 MOUSE	O54991 mus musculus
33	101	3.2	560	1 EFS_MOUSE	O64355 mus musculus

ALIGNMENTS

RESULT 1

ID	VMD2_HUMAN	STANDARD;	PRT;	585 AA.
AC	O76090; O75904;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bestrophin (Vitellogenesis macular dystrophy protein) (TUI5B).			
GN	VMD2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANTS BMD.			
RX	MEDLINE=98367043; PubMed=9700209;			
RA	Marquardt A., Stoehr H., Passmore L.A., Kraemer F., Rivera A.,			
RA	Weber B.H.P.;			
RT	"Mutations in a novel gene, VMD2, encoding a protein of unknown			
RT	properties cause juvenile-onset vitelliform macular dystrophy (Best's			
RT	disease).";			
RL	Hum. Mol. Genet. 7:1517-1525(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND VARIANTS BMD P-6; H-85; C-93; N-227 AND E-299.			
RX	MEDLINE=98324772; PubMed=96623195;			
RA	Petrushkin K., Koisti M.J., Bakall B., Li W., Xie G., Marknell T.,			
RA	Sandgren O., Forsman K., Holmgren G., Andreasson S., Vujic M.,			
RA	Bergsen A.A.B., McCarty-Dugan V., Figueroa D., Austin C.P.,			
RA	Metzker M.L., Caskey C.T., Wadelius C.;			
RT	"Identification of the gene responsible for Best macular dystrophy.;"			
RL	Nat. Genet. 19:241-247(1998).			
RN	[3]			
RP	VARIANTS BMD HIS-13; CYS-93; CYS-218; ASP-300; GLU-301 AND ILE-307.			
RX	MEDLINE=99265978; PubMed=10331951;			
RA	Caldwell G.M., Kakuk L.B., Griesinger I.B., Simpson S.A., Nowak N.J.,			
RA	Small K.W., Maunee I.H., Rosenfeld P.J., Sieving P.A., Shows T.B.,			
RA	Ayagari R.;			
RT	"Bestrophin gene mutations in patients with Best vitelliform macular			
RT	dystrophy.;"			
RL	Genomics 58:98-101(1999).			
RN	[4]			
RP	VARIANTS BMD V-10; V-82; C-92; H-96; S-135; C-218; S-218 AND K-293.			
RX	MEDLINE=99320852; PubMed=10394929;			
RA	Bakall B., Marknell T., Ingvas S., Koisti M.J., Sandgren O., Li W.,			
RA	Bergsen A.A.B., McCarty-Dugan V., Figueroa D., Austin C.P.,			
RT	"The mutation spectrum of the bestrophin protein -- functional			
RT	implications.;"			
RL	Hum. Genet. 104:383-389(1999).			
RN	[5]			
RP	VARIANTS BMD/BMD, VARIANT AVMD K-146, AND VARIANT BULL'S EYE Q-119.			
RX	MEDLINE=99381534; PubMed=10453731;			
RA	Allikmets R., Seddon J.M., Bernstein P.S., Hutchinson A., Atkinson A.,			
RA	Sharma S., Gerrard B., Li W., Metzker M.L., Wadelius C., Caskey C.T.,			
RA	Dean M., Petruskin K.;			
RT	"Evaluation of the Best disease gene in patients with age-related			
RT	macular degeneration and other maculopathies.;"			

O89853 ebola virus
Q03701 homo sapien
P03204 epstein-bar
P16952 streptococc
P48415 saccharomyc
Q98866 rhizobium l
Q98146 kapoei's sa
Q43125 arabidopsis
Q8yus anabaena sp
P02549 homo sapien
P16649 saccharomyc
P78357 homo sapien

34 100.5 3.2 577 1 VGP EBORS
35 100.5 3.2 998 1 CBF_HUMAN
36 100 3.2 992 1 EBN6_EBV
37 99.5 3.2 1500 1 SSP5_STRGN
38 99.5 3.2 2195 1 SC16_YEAST
39 99 3.2 309 1 Y716_RHILO
40 99 3.2 342 1 VG74_KSHV
41 98.5 3.2 681 1 CRY1_ARATH
42 98 3.1 307 1 Y787_ANASP
43 98 3.1 2418 1 SPCA_HUMAN
44 97.5 3.1 713 1 TUP1_YEAST
45 97.5 3.1 1384 1 CTA1_HUMAN

Hum. Genet. 104:449-453 (1999).

[6] VARIANTS BMD F-16; C-17; N-73; H-92; C-218; H-218; L-235 AND S-296.

RA MEDLINE=21138457; PubMed=11241846; Dureau P., Gogat K., Boutboul S., Pequignot M., Sternberg C., Dumur J., Mouillon M., Munier F.L., Schorderet D.F., Marsac C., Dufier J.L., Abitbol M., Identification of novel VMD2 gene mutations in patients with Best vitelliform macular dystrophy."

RT Hum. Mutat. 17:235-235 (2001).

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- DISEASE: DEFECTS IN VMD2 ARE THE CAUSE OF BEST MACULAR DYSTROPHY (BMD). ALSO KNOWN AS VITELLIFORM MACULAR DYSTROPHY TYPE 2. BMD IS AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY TYPICAL "EGG-YOLK" MACULAR LESIONS DUE TO ABNORMAL ACCUMULATION OF LIPOFUSCIN WITHIN AND BENEATH THE RETINAL PIGMENT EPITHELIUM CELLS. PROGRESSION OF THE DISEASE LEADS TO DESTRUCTION OF THE RETINAL PIGMENT EPITHELIUM AND VISION LOSS.

CC -!- DISEASE: DEFECTS IN VMD2 COULD BE THE CAUSE OF OTHER FORMS OF MACULOPATHY AS BULL'S EYE MACULOPATHY, AND ADULT VITELLIFORM MACULAR DEGENERATION (AVMD).

CC -!- DISEASE: IN RARE CASES, DEFECTS IN VMD2 MAY INCREASE SUSCEPTIBILITY TO AGE-RELATED MACULAR DEGENERATION (AMD).

CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

CC -!- DATABASE: NAME=VMD2 mutation database; WWW="http://www.uni-wuerzburg.de/humangenetics/vmd2.html".

CC -!- DATABASE: NAME=Mutations of the VMD2 gene; NOTE=Retina International's Scientific Newsletter; WWW="http://www.retina-international.com/sci-news/vmd2mut.htm".

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF073500; AAC64926.1; JOINED.

DR EMBL; AF073491; AAC64926.1; JOINED.

DR EMBL; AF073492; AAC64926.1; JOINED.

DR EMBL; AF073493; AAC64926.1; JOINED.

DR EMBL; AF073494; AAC64926.1; JOINED.

DR EMBL; AF073495; AAC64926.1; JOINED.

DR EMBL; AF073496; AAC64926.1; JOINED.

DR EMBL; AF073497; AAC64926.1; JOINED.

DR EMBL; AF073498; AAC64926.1; JOINED.

DR EMBL; AF073499; AAC64926.1; JOINED.

DR EMBL; AF057169; AAC64343.1; -.

DR EMBL; AF057170; AAC64344.1; -.

DR EMBL; AF073501; AAC33766.1; -.

DR Genew; HGNC:12703; VMD2.

DR MIM; 153700; -.

DR InterPro; IPR000615; Worm_fam_8.

DR Pfam; PF01062; DUF289; 1.

DR ProDom; PD002802; Worm_fam_8; 1.

KW Alternative splicing; Disease mutation; Polymorphism; Vision.

FT VARSPLIC 291 483

FT AEQLINPFGEEDDDFTNWIWVDRASQPRASFMGSTFNISLNKEE

FT MEPTDYNKPEQPYPPTAASQPRASFMGSTFNISLNKEE

FT MFPQPNQDEEDAHAGIIGRFLGSHDHPHPRANSRTKL

FT WPKRESLLHGLPKNHKAARKONVRGOEDHNKAKLKVADPK

FT SAPLYORPGVYSAPQTPSLPTMFFLEP -> GLSRALLG

FT WRGQRGHGQQLLETMRQCOERKVSVESSQAWRTVLP

FT TRAEAGLESLEPGRRRLWQSSSTSLERMMILRTGLST

FT GICRCPCWLMRRTCTGLGWSRTCTGISPSHSPTQLLPPS

FT SVEPPIWAPPST (IN ISOFORM 2).

FT T -> P (IN BMD AND AVMD).

FT /FTid=VAR_000830.

FT V -> A (IN BMD).

FT /FTid=VAR_000831.

FT V -> M (IN BMD).

FT

FT VARIANT 6 6

FT VARIANT 9 9

FT VARIANT 9 9

FT

FT /FTid=VAR_000832.

FT A -> T (IN BMD).

FT /FTid=VAR_000833.

FT A -> V (IN BMD).

FT /FTid=VAR_010468.

FT R -> H (IN BMD).

FT /FTid=VAR_010469.

FT S -> F (IN BMD).

FT /FTid=VAR_010470.

FT F -> C (IN BMD).

FT /FTid=VAR_010471.

FT L -> V (IN BMD).

FT /FTid=VAR_000834.

FT W -> C (IN BMD).

FT /FTid=VAR_000835.

FT R -> Q (IN BMD).

FT /FTid=VAR_000836.

FT R -> W (IN BMD).

FT /FTid=VAR_000837.

FT S -> R (IN BMD).

FT /FTid=VAR_000838.

FT Q -> L (IN BMD).

FT /FTid=VAR_000839.

FT L -> V.

FT /FTid=VAR_000840.

FT I -> N (IN BMD).

FT /FTid=VAR_010472.

FT L -> V (IN BMD).

FT /FTid=VAR_010473.

FT Y -> H (IN BMD).

FT /FTid=VAR_000841.

FT R -> C (IN BMD).

FT /FTid=VAR_010474.

FT R -> H (IN BMD).

FT /FTid=VAR_010475.

FT R -> S (IN BMD).

FT /FTid=VAR_000842.

FT W -> C (IN BMD).

FT /FTid=VAR_000843.

FT Q -> H (IN BMD).

FT /FTid=VAR_010476.

FT N -> K (IN BMD).

FT /FTid=VAR_000844.

FT L -> R (IN BMD).

FT /FTid=VAR_000845.

FT D -> E (IN BMD).

FT /FTid=VAR_000846.

FT E -> Q (IN BULL'S EYE MACULOPATHY).

FT /FTid=VAR_010477.

FT G -> S (IN BMD).

FT /FTid=VAR_010478.

FT R -> H (IN BMD).

FT /FTid=VAR_000847.

FT A -> K (IN AVMD).

FT /FTid=VAR_010479.

FT S -> N (IN BMD).

FT /FTid=VAR_000848.

FT T -> I (IN AMD; SPORADIC).

FT /FTid=VAR_010480.

FT R -> C (IN BMD).

FT /FTid=VAR_000849.

FT R -> H (IN BMD).

FT /FTid=VAR_010481.

FT R -> Q (IN BMD).

FT /FTid=VAR_000850.

FT R -> S (IN BMD).

FT /FTid=VAR_000851.

FT L -> M (IN BMD).

FT /FTid=VAR_000852.

FT Y -> C (IN BMD).

FT /FTid=VAR_000853.

FT Y -> N (IN BMD).

FT /FTid=VAR_000854.

FT

Query Match 99.7%; Score 3110; DB 1; Length 585;
 Best Local Similarity 99.7%; Pred. No. 4.7e-225;
 Matches 583; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLCYIIRFYRLALTEEQOL 60
 DB 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLCYIIRFYRLALTEEQOL 60
 QY 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEGKDEQ 120
 DB 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEGKDEQ 120
 QY 121 SLLRLRTLIYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 DB 121 GLLRLRTLIYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 QY 181 FWPPVWFANLSMKAWLGGRIIDPILQLSLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240
 DB 181 FWPPVWFANLSMKAWLGGRIIDPILQLSLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240
 QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 QY 301 DDDDPETNWIIDRNLOVSLAVDEMHDLPNEPDMYWNKPEPPPYTAASQFRASPM 360
 DB 301 DDDDPETNWIIDRNLOVSLAVDEMHDLPNEPDMYWNKPEPPPYTAASQFRASPM 360
 QY 361 GSTFNISLNK--EEMEFQPNQDEEDAHGIIIGFLGQSHDHPHPRANSRTKLLPKRESL 420
 DB 361 GSTFNISLNK--EEMEFQPNQDEEDAHGIIIGFLGQSHDHPHPRANSRTKLLPKRESL 420
 QY 421 LHEGLPKNHKAQKQNVRGQEDNKAWLKAVDAFKSGPLYQRPYYSAPOTPLSPPTMFPF 480
 DB 421 LHEGLPKNHKAQKQNVRGQEDNKAWLKAVDAFKSGPLYQRPYYSAPOTPLSPPTMFPF 480
 QY 481 LEPAPSCLKHVTGIDTKDKSLKTVSSGAKKSFELLSSDGLMEHPEVSVRRKTVFN 540
 DB 481 LEPAPSCLKHVTGIDTKDKSLKTVSSGAKKSFELLSSDGLMEHPEVSVRRKTVFN 540
 QY 541 LTDMPPEIPENHLKPELQSPNTNIHTTLKHMDPYWALENRDEAHS 585
 DB 541 LTDMPPEIPENHLKPELQSPNTNIHTTLKHMDPYWALENRDEAHS 585

RESULT 2
 Y022 CAEEL STANDARD; PRT; 632 AA.
 AC P34672;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK688.2 in chromosome III.
 GN ZK688.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodexinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins J., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showken R.,
 RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,

Wohlman P.;
 "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 CC 1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; LI6621; AAA28228.1; -
 DR PIR; S44917; S44917.
 DR WormPep; ZK688.2; CE00460.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 73829 MW; 0ABDD1755EF11642 CRC64;
 Query Match 24.1%; Score 752; DB 1; Length 632;
 Best Local Similarity 35.9%; Pred. No. 1.2e-48;
 Matches 152; Conservative 94; Mismatches 163; Indels 14; Gaps 4;

QY 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLCYIIRFYRLALTEEQOL 60
 DB 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLCYIIRFYRLALTEEQOL 60
 QY 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEGKDEQ 120
 DB 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEGKDEQ 120
 QY 121 SLLRLRTLIYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 DB 121 SLLRLRTLIYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 QY 181 FWPPVWFANLSMKAWLGGRIIDPILQLSLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240
 DB 181 FWPPVWFANLSMKAWLGGRIIDPILQLSLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240
 QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 QY 301 DDDDPETNWIIDRNLOVSLAVDEMHDLPNEPDMYWNKPEPPPYTAASQFRASPM 360
 DB 301 DDDDPETNWIIDRNLOVSLAVDEMHDLPNEPDMYWNKPEPPPYTAASQFRASPM 360
 QY 361 GSTFNISLNK--EEMEFQPNQDEEDAHGIIIGFLGQSHDHPHPRANSRTKLLPKRESL 420
 DB 361 GSTFNISLNK--EEMEFQPNQDEEDAHGIIIGFLGQSHDHPHPRANSRTKLLPKRESL 420
 QY 421 LHEGLPKNHKAQKQNVRGQEDNKAWLKAVDAFKSGPLYQRPYYSAPOTPLSPPTMFPF 480
 DB 421 LHEGLPKNHKAQKQNVRGQEDNKAWLKAVDAFKSGPLYQRPYYSAPOTPLSPPTMFPF 480
 QY 481 LEPAPSCLKHVTGIDTKDKSLKTVSSGAKKSFELLSSDGLMEHPEVSVRRKTVFN 540
 DB 481 LEPAPSCLKHVTGIDTKDKSLKTVSSGAKKSFELLSSDGLMEHPEVSVRRKTVFN 540
 QY 541 LTDMPPEIPENHLKPELQSPNTNIHTTLKHMDPYWALENRDEAHS 585
 DB 541 LTDMPPEIPENHLKPELQSPNTNIHTTLKHMDPYWALENRDEAHS 585

RESULT 3
 YS63 CAEEL STANDARD; PRT; 405 AA.
 ID YS63 CAEEL STANDARD; PRT; 405 AA.
 AC Q09379;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK675.3 in chromosome II.
 GN ZK675.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodexinae; Caenorhabditis.
 OX NCBI_TaxID=6239;


```

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z81074; CAB03043.1; -
DR WormPep; F32B6.9; CE09864.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 413 AA; 48965 MW; A66E69A83C78790B CRC64;
-----
Query Match 21.9%; Score 682; DB 1; Length 413;
Best Local Similarity 39.6%; Pred. No. 1.1e-43;
Matches 141; Conservative 67; Mismatches 126; Indels 22; Gaps 7;
QY 1 MTITYTSOVANARLGSFRLLLCWGSIYKLLYGBFLIFLLCYIIRFYRLAL-----T 55
Db 1 MTISYS-----GNVIRILLRWKGSIWRTAKELLIIILYYSVRVYFKGIDILDDDD 52
QY 56 EQQL-----MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRMNQYENLPWDRMLSLVS 111
Db 53 EDRRLKMRMEPTFCQDCDSYTRLLPLTFLGFGYVNVVWRWQFETLYWPELILSVLC 112
QY 112 GFVSGKDFQSLRLRLTLRYANLGNVILRSVSTAVYKRPSPQAHLVQAGFWTPAEHKQL 171
Db 113 TVLHQHDEKSKRRRTIARYLANALAWRDISSKIRLRFPSVHSLIESGLLTKYEQIL 172
QY 172 EKLSLPH-NMFWVPVWVFANLSMKAWLGGRIIDPILLOSLNEMNTLRTQCCHLYAYDWT 230
Db 173 EAMAEKSSRWITPLHWIQLIMRQVEEHKPTASLNFQVGLRIFRQSLRKLVSNDV 232
QY 231 SIPLVYTVQVTVAVYSFFLTCLVGRQFLNPAKAYEGHEDLVVPVFTFLQFFFYVGLVKV 290
Db 233 CVPLVYTVQVTVAVYSFFLTCLVGRQFLNPAKAYEGHEDLVVPVFTFLQFFFYVGLVKV 291
QY 291 AEQLINFGEDDDDFETNWIIVDRNLQVSLAVDEMHDQ-LPRME--DDMYNKNKPS 343
Db 292 QDLMRPPEGLDDDIELNYILDRNVRISFAIVNQLQESPIPDFESNDKLMHEMHP 347
-----
RESULT 6
QY03 CAEEL STANDARD; PRT; 450 AA.
AC Q17528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein B0564.3 in chromosome IV.
GN B0564.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lightning J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73422; CAA97765.1; -
DR WormPep; B0564.3; CE05177.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 450 AA; 53275 MW; A0PED9A476166AD7 CRC64;
-----
Query Match 21.8%; Score 679; DB 1; Length 450;
Best Local Similarity 32.9%; Pred. No. 2.1e-43;
Matches 155; Conservative 80; Mismatches 184; Indels 52; Gaps 9;
QY 1 MTITYTSOVANARLGSFRLLLCWGSIYKLLYGBFLIFLLCYIIRFYRLALTEEQQL 60
Db 1 MTINHKEIMTSHPTWTFLLFKWGSIKVAVMETIIFLICYGIIISVIYKTAMGESSQR 60
QY 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRMNQYENLPWDRMLSLVSGFVEGKDEQ 120
Db 61 TTESLVRYFDKLSYIPLSEFVLGFFVTVVNRWTKLYQTIGFDNVGLMANCYIRGATEK 120
QY 121 SLLRLTLRYANLGNVILRSVSTAVYKRPSPQAHLVQAGFWTPAEHKOLEKLSLPHN- 179
Db 121 ARIYERNIMRYCELQVILVFRDMSWTRRRFPTMETVVAAGFMNKHELELYNSYDTKYNS 180
QY 180 ----MFWVPVWVFANLSMKAWLGGRIIDPILLOSLNEMNTLRTQCCHLYAYDWTISPLV 235
Db 181 KLGTWIPANWALCMTYKARKDGYIESDYFKAQMEGEIRTWRTNIEWVCNVDWVPLPM 240
QY 236 YQVTVAVYSFFLTCLVGRQFLNPAKAYEGHEDLVVPVFTFLQFFFYVGLVKVABOLI 295
Db 241 YPOLVCLANLYFLYSIIARQ-LVIEKHQWDEVVDVYFPVMTFLOIFFYGLKVIDVML 299
QY 296 NPFGEEDDDDFETNWIIVDRNLQVSLAVDEMHDQ-LPRMEPDMYKNKPEPPPYTAASAQFR 355
Db 300 NPFGEEDDDDFETNWIIVDRNLQVSLAVDEMHDQ-LPRMEPDMYKNKPEPPPYTAASAQFR 358
QY 356 RASFGMSTFNISLN-----KSEMEFPQNOEDEDEDAHAGIIGFLGL 396
Db 359 NHHYHGSVSEVLEOKGNAPVMMPHSGSAANLRMMSPKSVDEDEKONAFSM----- 412
QY 397 QSHDHPHPRANSRTKLLMPKRESLLHLEGLPKNHKAQKQNVRGQEDNKAWKL 447
Db 413 -SHDD-ARRN---W--REVSLDSSF-----LADLNENKWKI 443
-----
RESULT 7
QY04 CAEEL STANDARD; PRT; 523 AA.
AC Q17529;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein B0564.4 in chromosome IV.
GN B0564.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lightning J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: Z73422; CAA97766.1; -
 CC WormPep; B0564.4; CE05178.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 53 AA; 61725 MW; EC37P07253E5P9A6 CRC64;

Query Match 21.5%; Score 669.5; DB 1; Length 523;
 Best Local Similarity 36.5%; Pred. No. 1.3e-42;
 Matches 142; Conservative 73; Mismatches 165; Indels 9; Gaps 4;
 QY 1 MTITYTSOVANARLGSFRLLLCMRGSYKLYGFLFLPLCYIIRFYRLALTEEQ 60
 DB 1 MTINHKETSHTWKFFVLLFRWKSILWKALYMETIIFLCYGLISVYRTAMSEPS 60
 QY 61 MFEKLTLYCDSYIOLIPISFVLGFTVTLVTRWNNQYENLWPDRLMSLVSGFVGEKDEQ 120
 DB 61 TFSVIRYCDKRLSFIPLEFVLGFTVIVDRWTKLWRTVGFIDVCLLANLYVGTSEK 120
 QY 121 SRLRLRLIRVANLGNVILRSVAVYKRPSPAOHLVQAGPMTPAEHKQLEKLSLPHN- 179
 DB 121 AIIYRRNIARYCALTLVFRDVSMTTRRPPTMETVVAAGFMSKDELGLYNSYTTKNS 180
 QY 180 ---MFVVPWVPANLSKMWLGRIRDPILLQSLNEMNTLRTQCGHLYAVDMISIP 235
 DB 181 RLKXKXIPANWALCMYKARQGYSDYFKAQMEGBIRTWRTNIEWCNVDWVPLPLM 240
 QY 236 YTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVMLKVAEQL 295
 DB 241 YPQLVCLAVNLVFLVSIARQ-LVIEKHWDVDEVDPVFTFLQFFYVGVMLKVAEQL 299
 QY 296 NPFGEDEDDFTFNWIVDRNLQVSLAVDEMHDLPMEPDPMYKPEQPPYTAASQPR 355
 DB 300 NPFGEDEDDFTFNWIVDRNLQVSLAVDEMHDLPMEPDPMYKPEQPPYTAASQPR 358
 QY 356 RASPMGTFNLSLKE--EMEFQPNQDE 382
 DB 359 NYHVGSTSEVHLSEKSSVRMIPHSQSE 387

RESULT 8
 YKTS CAEEL
 ID -YKTS CAEEL STANDARD; PRT; 459 AA.
 AC P34319;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C07A9.8 in chromosome III.
 OS C07A9.8.
 GN Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Cooper J., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Favello A., Fraser A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Hillier L., Hillier L., Jier M.,
 RA Pulten L., Gardner A., Green P., Hawkes T., Hillier L., Laister N.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
 RA Sime M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulten J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wodman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";
 RL Nature 368:32-38 (1994).
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Z29094; CAA82342.1; -
 DR PIR; S40708; S40708.
 DR WormPep; C07A9.8; CE00497.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 459 AA; 53570 MW; C2D8P69E4078BF9F CRC64;

Query Match 21.0%; Score 656; DB 1; Length 459;
 Best Local Similarity 32.3%; Pred. No. 1.1e-41;
 Matches 129; Conservative 91; Mismatches 151; Indels 28; Gaps 5;
 QY 1 MTITYTSOVANARLGSFRLLLCMRGSYKLYGFLFLPLCYIIRFYRLALTEEQ- 59
 DB 44 LSYNTYDLATSKSLMIVRMIFKWRGSLVQAVYKELIVWICAYSLSVIYRFPALTRSOKE 103
 QY 60 -----LMPEKLTLYCDSYIOLIPISFVLGFTVTLVTRWNNQYENLWPDRLMSLVSGFV 114
 DB 104 QNKEIIIFERECEYCDARMGYLPLNVLGFFCNIIIRRWLKLTYSLGNIDNIALFVSAYV 163
 QY 115 EGKDSQSLLARTLIRVANLGNVILRSVAVYKRPSPAOHLVQAGPMTPAEHKQLEK 174
 DB 164 RTDDRAQIRNIIRYICVISQCLVFRDTHVGVRRFRFFLEAVQAGIMLPHELEKFN 223
 QY 175 SLPHNMFVVPWVPANLSKMWLGRIRDPILLQSLNEMNTLRTQCGHLYAVDMISIP 234
 DB 224 KSRYQKYVSPFNWALELLNVAKTEKSIDGNARNAIAQESKFRSALTTVSMYDWP 283
 QY 235 YTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVMLKVAEQL 294
 DB 284 MYPQLVNNVAVHTYFPCITFTRQFFISADAHNKTEVDLYIPFMTIIEFIYMGWLKVAMEL 343
 QY 295 INPGEDEDDFTFNWIVDRNLQVSLAVDEMHDLPMEPDPMYKPEQPP- 346
 DB 344 LNPFGEDADDFTCNLLIDRNLAIGLTSVDDAYDQLPEVKPDPVFTGSGVKPLSDSDTRSLK 403
 QY 347 -YTAASQPRASPMGTFNLSLKEEMEF-----QPNQ 379
 DB 404 YHFGSAAQMEISY-----LKKEENKMAAGKKPK 434

RESULT 9
 YXAK CAEEL
 ID -YXAK CAEEL STANDARD; PRT; 513 AA.
 AC Q21973;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein R13.3 in chromosome IV.
 GN R13.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RC Matthews P.;
 RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

CC

--- THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isg-gib.ch/announce/ or send an email to licenses@isg-gib.ch).

CC

EMBL; Z73105; CAA97442.1; --

DR WormPep; R13_3; CE06320.

DR InterPro; IPR000615; Worm_fam_8.

DR Pfam; PF01062; DUF289; 1.

DR ProDom; PD02802; Worm_fam_8; 1.

KW Hypothetical protein.

SQ SEQUENCE 513 AA; 58919 MW; 666CB054AD179489 CRC64;

Query Match 20.7%; Score 646.5; DB 1; Length 513;
Best Local Similarity 30.3%; Pred. No. 6.8e-41;
Matches 160; Conservative .97; Mismatches 200; Indels 71; Gaps 13;

Qy 1 MTIYTSQVANARLGSPSRLLLRGSGTYKLYGEFLIFLLCYVIRPIYR--LAITEEQ 58
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 59 QLMPEKLTLYCDSYIQIPISFVLGFVVYLVTVRMMNQENLPWRMLSLVGFGVEGD 118
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 61 RKIWDNPAALFDQNMDFIPLTFMGLGFVTIVRRNDIPANLGHWENTAITVANYTRGD 120
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 119 EQSRLLRRTLIRYANLGNVILRSVSNAVYKRFPSPAHLVOAGFMTPAEHKLEKLSLPH 178
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 121 DRTMRIRNRVIRYMVLAQLVFRDCSIQVRKRPFTMESIVSAGSFQCLGSSTATE----- 175
Qy 179 NMFWVPVMWANLSMKAWLGGRTRDPILLQSLLNEMNTLTQCCHLYADWTISPIVYTQ 238
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 176 -----YYWGSLGVLDARAEGKIADLLMEIGKHIEFRKMALLSNYDMVPIPLAYPQ 229
Qy 239 VTVAVYSFPLTCLVGRQFL-----NPKAYPGHELDLVYP-VPTLQRFFYVGMWK 289
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 230 VFPLAVRSYFPMALIARQSVLLDCKEPEQSIIYP-----TVFVMSILQFIYVGMWK 283
Qy 290 VASQLINPFGEDDDDFFTNIWIDRNLOVSLVAVDEMHDQLPRMEPDYMNKPEPPPYTA 349
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 284 VAESMINPLGEDDDDDPCYNLLDNLMIGLCIVDDNVNRTFSVEKDFAW-CADVEPLYSV 342
Qy 350 ASAQFRASPWGS--TFNISLNKEEMEPQNQDEDAHGIIGRFLGLQSHDHHPRA- 406
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 343 ETAMI PKNPOIGSAANYDVKVDEEYMMMPHMDDVD-----LDFESTNNLIIPRKT 393
Qy 407 -----NSRTKLLPKRESLLHGLEGPLKNHKAANKVNGQEDNKAKLKAVDAFKSG- 456
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 394 FSVISTQRPGSGRASLASRK-R-SMWFDL-RGRIAKKKQRSNMFNQS----- 438
Qy 457 PLYQRPGYSAPQT-----LSPTMFPPLEPSAPSKLHSVTGITDKDK 500
Db |||:::||:::||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Qy 439--VSQASHVFESQAPSEINSLTLEMTPARKKSSTGKLSGSMVAEEQH 485

RESULT 10
ID YNX4_CABEL
AC YNX4_CABEL STANDARD; PRT; 456 AA.
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DE Hypothetical protein T20G5.4 in chromosome III.
OS T20G5.4.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RF SEQUENCE FROM N.A.

```

15-JUN-2002 (Rel. 41, Last annotation update)
Hypothetical protein T19C3.1 in chromosome III.
T19C3.1.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
SEQUENCE FROM N.A.
STRAIN=Bristol N2;
Fulton L.;
Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[2]
REVISIONS.
Waterston R.;
Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U28412; AAL56625.1; .
WormPep; T19C3.1; CE30182.
InterPro; IPR000615; Worm_fam_8.
Pfam; PF01062; DUF289; 1.
ProDom; PD002802; Worm_fam_8; 1.
Hypothetical protein.
KW
SEQUENCE 501 AA; 57930 MW; 1C56C4C7070FC62 CRC64;
Query Match
Best Local Similarity 20.0%; Score 622.5; DB 1; Length 501;
Matches 149; Conservative 84; Mismatches 175; Indels 75; Gaps 8;
QY 1 MTITTSQVANARLGSFSLLLCWGSIYKLLYGFLFLCYIIRF----- 50
DB 23 MTVSQLDVSSGNPLFLFLGRGSIWKSIVGDLFWLLFYAIYAFYAFSKQLQT 82
QY 61 MPEKLTLYCDYSIQIPIPSIFVLGFFVYVTVVTRMNOYENLPPDRLMSLVSGFVGKDEQ 120
DB 83 VPEISIHTRDKMYLPLTFMLGFFVTVTFWRWSALNVMPTESVALSVALLPKGKRE 142
QY 121 SLLRLRTLYRANLGNVILRSVSTAVYKRFPSAQLHVOAGFWTPAEHKQLEKLSLPHNM 180
DB 143 DLTRTRAIIRYVVLHQILVFRDISMKVRFRFPLTKYVDVAGFMRQBELDLVESVNOESSQ 202
QY 181 -FWPMWPFANLSMKAWLGGRIKRDPILOSLNEMNTLRQCGHLYAYDWISIPLYVTQV 239
DB 203 TWVFPINWANSALVAHQKLDIQPTAFNNVIFAIKEFRVAMETLIKEDAIPIPIAYPOV 262
QY 240 VTVAVYFFELTCLVGRQFL- NPAXVPGHELDLVVPVFTFLQFFVYVGHMLKVAEQLINFP 298
DB 263 VFLAVRVYFAICLVSRQFLSDMSKST--QMDVPVPIMTVLEFIFVIGMKVAEVLNPL 320
QY 299 GEDDDDDPETNWIIVDRNLQVSLAVDEMHDQLPRMEPDWYN----- 339
DB 321 GEDDDDDFEVNSIIDNNISRGMAIVDTTHGVHPDLVDVDFSDPNLYPAYSENSQIPNLITG 380
QY 340 -----KPEPPPYTAASAQF-----RRASPMGST----- 363
DB 381 SAAKVLAAPTDEKIVRVNPDAPPTSERGSMFTRNAYSLNRKISIASNNLESFPOE 440
QY 364 --FNIS-----LNKEMEPQNOEDEDAHAGIIGRFLGQSHDHPHPRANSTKLLWPK 416
DB 441 RKFNLSMPAGMLNKSTQPDRTMETVSEEH-----PSHFYRGDRVHSSDSVIPSP 491
QY 417 RES 419
DB 492 RRS 494
-----
RESULT 12
YCEL CAEEL STANDARD; PRT; 420 AA.
AC Q94175;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C43G2.4 in chromosome IV.
C43G2.4.
GN Caenorhabditis elegans.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton B.; Wohldmann P.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U70848; AAB09111.1; .
WormPep; C43G2.4; CE08693.
InterPro; IPR000615; Worm_fam_8.
Pfam; PF01062; DUF289; 1.
ProDom; PD002802; Worm_fam_8; 1.
Hypothetical protein.
KW
SEQUENCE 420 AA; 50223 MW; 5078D15E1D414A26 CRC64;
Query Match
Best Local Similarity 19.9%; Score 622; DB 1; Length 420;
Matches 151; Conservative 82; Mismatches 140; Indels 104; Gaps 16;
QY 1 MTITTSQVANARLGSFSLLLCWGSIYKLLYGFLFLCYIIRF----- 48
DB 1 MTISYS-----GNFRLDLRWKGSWKSVRLEFLFLFYFIRFSAHPFYNTDPT 52
QY 49 -----IYRLALTEEQLMFEKLTLYCDYSIQIPIPSIFVLGFFVYVTVVTRMNOYENLP 101
DB 53 DSKGYRKIFKVMCNFHE-----YTKMPLTLFLGFFYVSVNVRWVRQFETLR 100
QY 102 WPDRLMSLVSGFVGKDEKQSRLLRRLTYRANLGNVILRSVSTAVYKRFPSAQLHVOAG 161
DB 101 WPEDFLSILCLLLPSK--ESRPARHQIARYLNLTSALAWRDVSTKIRLRFPSLRNIIDAG 158
QY 162 FMTPAEHKQLEKLSLPHNMFW--PWVWFANL-SMKAWLGGRIKRDPI-LLOQL----- 211
DB 159 LUTEKEYEKLDQINWSKAVILMLPVYVTKNLCETSPSGIRLWTLFHLWVQQLIDAEITA 218
QY 212 -----NEMNTLRQCGHLYAYDWISIPLYVTQVTVVAVYVFFLTCVLRGQFLN 259
DB 219 GRGSVNYVSATNELKAYRISPRLLYCHDWCVPLVYTVQAALATYSYFFFLGFRQDLN 278
QY 260 PAKAYPGHELDLVVPVFTFLQFFVYVGHMLKVAEQLINPGEDEDDPETNWIIVDRNLQVSL 319
DB 279 HDDFY---SLDAFFFLFTVQVFLFFVGVFKVQDLMRPFGLDDDDFELSIIIDRNIVTSF 335
QY 320 LAVDEMH-QDLPRMEPDWYNKPEPPPYTAASAQFRRASFMSGTFNLSINKEEMEFQPN 378
DB 336 TIVDSLQDDDDPPKPFEDVFKKHNEQ-----QMHQSMFL-PRVPTSLKWRIDLSRN 387
QY 379 QBEDEDAHAGIIGRFLGQSHDHPHPRANSTKLLWPKRESLLHEGLPKKHKAQKN 435
DB 388 -----AH-----KHPPKLTYLEM-----KNQD-----PEYKRNKKN 415

```



```
RESULT 13
YV6L CAEEL
ID YV6L CAEEL STANDARD; PRT; 400 AA.
AC Q19978;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein F3268.4 in chromosome V.
GN F3268.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 272509; CAA96648.2; -
DR WormPep; F3268.4; CE23700.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 400 AA; 46750 MW; B1E8ABEC3B862E84 CRC64;
Query Match 18.8%; Score 585; DB 1; Length 400;
Best Local Similarity 31.5%; Pred. No. 1.9e-36;
Matches 140; Conservative 83; Mismatches 175; Indels 46; Gaps 11;
QY 1 MTITYTSQVANARLGSRLLLCWGRSIVKLYGFLIFLLCYIIRFYRLALTEEQOL 60
DB 1 MTISYDEB-----FSSLMLRWGSIWKAVLKDILGFIYIYIVLAFQVLLDEKGE 52
QY 61 MFEKLTLYCDYSYIQLIPISFVLGFFYVTLVTRWNNQYENLPWDRMLSLVSGFVEQDEQ 120
DB 53 YFTGWIMWCETGAQYIPLSFLGFPVSLIVARWMEQFNCISWPDQWIMVSAQLPG--NE 110
QY 121 SRLARRTLIRYANLGNVLILRSVSTAVYKRPFAOHLVOAGPMTPAEHKQLEKLSLPHN 180
DB 111 MNVVRQTIARWSSQAAIAWAGSVSKTLKRPFTERRHVMASKLMTBEEYDLYNMTDAPHGK 170
QY 181 FWVFWVWVFWANLGMKAWLGGRIIDPILLOSLLNEMNTLRTOCHLYAYDWISIPLVYTVQV 240
DB 171 WFIILMTVNLVKKQKQG--IIDSQMDMLLKQVYSYRDGFAMLPYDWMIKPLVYTVQV 229
QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVMLKVAEQLINPFE 300
DB 230 AIATYGVFFICLIGKQPKLDQRSME-KEITILPFIPTTFQMLFYLGWLKVGQ-----PSI 283
QY 301 DDDDFETNIVDRNLQVSLAVDEKHQDLPRMEPDMTNKPPEOPPYTAASQAQFRASFM 360
DB 284 RKKTSELNLYLDRNTATAHMAWSELSDQLPSIGAPV-----PAVPHTRASFKIQDVIPK 338
QY 361 GSTFNISLNKEMBE-FQPNQDEDEAHAGIIGRFLGLQSHDHPHPRANSRTKLLWPKEES 419
DB 339 SHLAGPKLSEAMKLIKP-----ED-----LEEHE-----RLMEETKVTNRRQLG 378
QY 420 LHEGILPKQNHKAAKQNVRGQEDNK 443
DB 379 TLVRAL---EKSRTNATINEDDE 399
RESULT 14
YAVK CAEEL
ID YAVK CAEEL STANDARD; PRT; 884 AA.
AC Q17851;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C09B9.3 in chromosome IV.
GN C09B9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Bradshaw H.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U50069; AAB37559.2; -
DR WormPep; C09B9.3; CE29571.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 2.
DR ProDom; PD002802; Worm_fam_8; 3.
DR KW Hypothetical protein; Transmembrane; Repeat.
FT TRANSMEM 265 285 POTENTIAL.
FT TRANSMEM 338 418 POTENTIAL.
SQ SEQUENCE 884 AA; 102851 MW; 8DB83F9699B8FAID CRC64;
Query Match 16.1%; Score 503.5; DB 1; Length 884;
Best Local Similarity 32.7%; Pred. No. 7.1e-30;
Matches 105; Conservative 70; Mismatches 139; Indels 7; Gaps 4;
QY 1 MTITYTSQVANARLGSRLLLCWGRSIVKLYGFLIFLLCYIIRFYRLALTEEQOL 60
DB 1 MTISYTDVATESYFGFPKVLFRWKGSKVLHRELFMWLVLYTVLAIYR-TLDEERKK 59
QY 61 MFEKLTLYCDYSYIQLIP--ISFVLGFFYVTLVTRWNNQYENLPWDRMLSLVSGFVEGKD 118
DB 60 IFRS----NIEHFINFEPSEILTFMLSFFVTIVQWRNNVFTNMGFTIENAAAYASVSEPMK-NG 115
QY 119 EQSILLRRTLIRYANLGNVLILRSVSTAVYKRPFAOHLVOAGPMTPAEHKQLEKLSLPH 178
DB 116 EDVRAQRTVIRYLVASQILVNRISIKALRRFFNYESIVTAGFTKEESTIIONTDLSY 175
QY 179 NMFWVWVWVFWANLGMKAWLGGRIIDPILLOSLLNEMNTLRTOCHLYAYDWISIPLVYTVQ 238
DB 176 DSSCVPIRWAIOVLRHQVRSNGFFSHSVYRATWKEVSDFETHLSRVKRVQWVPIFLAYPQ 235
QY 239 VTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVMLKVAEQLINPFF 298
DB 236 VIFFAVRLYFVICAPAKQYFDLDDDDARYVIHYIYFPFIVTVFQFICLMGWLKVAEALLNPL 295
```



```

QY 299 GEDDDDFETNWIIVDRNLQVSL 319
Db 296 GEDDDDFEVNFLIDSNIYTM 316

RESULT 15
YHS4_CAEEL STANDARD; PRT; 602 AA.
AC O18303;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK849.4 in chromosome I.
GN ZK849.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
REVIEWS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z82095; CAB05027.2; -
DR WormPep; ZK849.4; CE25696.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 2.
DR Hypothetical protein.
KW SEQUENCE 602 AA; 67607 MW; 458AB78802BD6B3A CRC64;
SQ
Query Match 15.2%; Score 474.5; DB 1; Length 602;
Best Local Similarity 29.5%; Pred. No. 6.2e-28;
Matches 135; Conservative 65; Mismatches 147; Indels 111; Gaps 12;

QY 1 MTITVTSQVANARLGSRLLCWRGSIYKLYGBFLIFLCYIIRFYRLAL-----T 55
Db 14 MTVTYNRAVSTESIHNFVSIWYHSGSLVKSIVKEYAIWLVLYHLFIYRVVMPFGWA 73
QY 56 BEQQLMFEKLYCDSYQLQIPISVLGFYTVLVVTRWNNQVENLPWDRMLSLVSGFVE 115
Db 74 DYCKRVIENTWTPHQD---MTIPLELLGFFVTVIDRWKAFQNIPIYIEICAFVAAAIP 130
QY 116 G-----KDEQSR----- 122
Db 131 GRIRIINKQNEVRAPGQDLDPVAVENHPVQRPGEAPAAALPMGAPERPAVALPMP 190
QY 123 -----LLRRTLIRYANLGNVLILRSVSTAVYKRPFSQAHLVQAGFMT 164
Db 191 GAPEGPAAIQOPSDVVKLTARRTIIRYLVLSQLILLFREISTVKKRFVDLCLVDSKELT 250
QY 165 PAEHKOLEK-LSLPH-NMFVWPWW-FANLS-----MKAWLGGRIKDPILLQSL 211
Db 251 DEELKILSKVKCHDYSIFLPINWAFSILQEHKVKSNPEFINAW--NVIRDWQVKLSLL 308
QY 212 NEMNTLRQCGLXAYDWISIPLVYTVVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDL 271
Db 309 RN-----GDFIPIPLAYQAVFLAIRFYFLVCLFTROHLMDDK---KSIDY 352

```

```

QY 272 VVPVFTLQFFYVGLKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLIIVDEMHDLP 331
Db 353 YFPLMTSLQFIFVGMKVAEILLNPMGEDDDDDFELNNIIDKLYIGLAIVTECGKHPE 412
QY 332 MEPDMYNNKPSPPPYTAASAQFRDASPMWGSTFNISLN 369
Db 413 IVKDTIGKDCLPFFYPQNDNERNRA-LVGSTKNINLS 449

```

Search completed: July 10, 2003, 12:24:43
Job time : 12.6817 secs